

## CHEMICAL INHIBITORS OF MISMATCH REPAIR

### TECHNICAL FIELD OF THE INVENTION

The invention is related to the area of mutagenesis. In particular it is related to the field of blocking specific DNA repair processes.

### BACKGROUND OF THE INVENTION

Mismatch repair (MMR) is a conserved DNA repair process that is involved in post-replicative repair of mutated DNA sequences that occurs after genome replication. The process involves a group of gene products, including the *mutS* homologs *GTBP*, *hMSH2*, and *hMSH3* and the *mutL* homologs *hMLH1*, *hPMS1*, and *hPMS2* (Bronner, C.E. *et al.* (1994) *Nature* 368:258-261; Papadopoulos, N. *et al.* (1994) *Science* 263:1625-1629; Leach, F.S. *et al.* (1993) *Cell* 75:1215-1225; Nicolaides, N.C. *et al.* (1994) *Nature* 371:75-80) that work in concert to correct mispaired mono-, di-, and tri-nucleotides, point mutations, and to monitor for correct homologous recombination. Germline mutations in any of the genes involved in this process results in global point mutations, and instability of mono, di and tri-nucleotide repeats (a feature referred to as microsatellite instability (MI)), throughout the genome of the host cell. In man, genetic defects in MMR results in the predisposition to hereditary nonpolyposis colon cancer, a disease in which tumors retain a diploid genome but have widespread MI (Bronner, C.E. *et al.* (1994) *Nature* 368:258-261; Papadopoulos, N. *et al.* (1994) *Science* 263:1625-1629; Leach, F.S. *et al.* (1993) *Cell* 75:1215-1225; Nicolaides, N.C. *et al.* (1994) *Nature* 371:75-80; Harfe B.D., and S. Jinks-Robertson (2000) *An. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). Though the mutator defect that arises from MMR deficiency can affect any DNA sequence, microsatellite sequences are particularly sensitive to MMR abnormalities (Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069). Microsatellite instability is therefore a useful indicator of defective MMR. In addition to its occurrence in virtually all tumors arising in HNPCC patients, MI is found in a small fraction of sporadic tumors with distinctive molecular and

phenotypic properties that is due to defective MMR (Perucho, M. (1996) *Biol. Chem.* 377:675-684).

MMR deficiency leads to a wide spectrum of mutations (point mutations, insertions, deletions, recombination, etc.) that can occur throughout the genome of a host cell. This effect has been found to occur across a diverse array of organisms ranging from but not limited to unicellular microbes, such as bacteria and yeast, to more complex organisms such as *Drosophila* and mammals, including mice and humans (Harfe B.D., and S. Jinks-Robertson (2000) *An. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). The ability to block MMR in a normal host cell or organism can result in the generation of genetically altered offspring or sibling cells that have desirable output traits for applications such as but not limited to agriculture, pharmaceutical, chemical manufacturing and specialty goods. A chemical method that can block the MMR process is beneficial for generating genetically altered hosts with commercially valuable output traits. A chemical strategy for blocking MMR *in vivo* offers a great advantage over a recombinant approach for producing genetically altered host organisms. One advantage is that a chemical approach bypasses the need for introducing foreign DNA into a host, resulting in a rapid approach for inactivating MMR and generating genetically diverse offspring or sib cells. Moreover, a chemical process is highly regulated in that once a host organism with a desired output trait is generated, the chemical is removed from the host and its MMR process would be restored, thus fixing the genetic alteration in subsequent generations. The invention described herein is directed to the discovery of small molecules that are capable of blocking MMR, thus resulting in host organisms with MI, a hallmark of MMR deficiency (Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Perucho, M. (1996) *Biol. Chem.* 377:675-684; Wheeler, J.M. *et al.* (2000) *J. Med. Genet.* 37:588-592; Hoang, J.M. *et al.* (1997) *Cancer Res.* 57:300-303). Moreover, host organisms exhibiting MI are then selected for to identify subtypes with new output traits, such as but not limited to mutant nucleic acid molecules, polypeptides, biochemicals, physical appearance at the microscopic and/or macroscopic level, or phenotypic alterations in a whole organism. In addition, the ability to develop MMR defective host cells by a chemical agent provides a valuable method for creating genetically altered cell hosts for product development. The invention described herein is directed to the

creation of genetically altered cell hosts via the blockade of MMR using chemical agents *in vivo*.

The advantages of the present invention are further described in the examples and figures described within this document.

### SUMMARY OF THE INVENTION

The invention provides methods for rendering cells hypermutable by blocking MMR activity with chemical agents.

The invention also provides genetically altered cell lines which have mutations introduced through interruption of mismatch repair.

The invention further provides methods to produce an enhanced rate of genetic hypermutation in a cell.

The invention encompasses methods of mutating a gene of interest in a cell, methods of creating cells with new phenotypes, and methods of creating cells with new phenotypes and a stable genome.

The invention also provides methods of creating genetically altered whole organisms and methods of creating whole organisms with new phenotypes.

These and other objects of the invention are provided by one or more of the embodiments described below.

In one embodiment of the invention, a method for screening chemical compounds that block mismatch repair (MMR) is provided. An MMR-sensitive reporter gene containing an out-of-frame polynucleotide repeat in its coding region is introduced into an MMR proficient cell. The cell is grown in the presence of chemicals. Chemicals that alter the genetic structure of the polynucleotide repeat yield a biologically active reporter gene product. Chemicals that disrupt the polynucleotide repeat are identified as MMR blocking agents.

In another embodiment of the invention, an isolated MMR blocking chemical is provided. The chemical can block MMR of a host cell, yielding a cell that exhibits an enhanced rate of hypermutation.

In another embodiment of the invention, a method is provided for introducing a mutation into a gene of interest. A chemical that blocks mismatch repair is added to the

culture of a cell line. The cells become hypermutable as a result of the introduction of the chemical. The cell further comprises a gene of interest. The cell is cultured and tested to determine whether the gene of interest harbors a mutation.

In another embodiment of the invention, a method is provided for producing new phenotypes of a cell. A chemical that blocks mismatch repair is added to a cell culture. The cell becomes hypermutable as a result of the introduction of the chemical. The cell is cultured and tested for the expression of new phenotypes.

In another embodiment of the invention, a method is provided for restoring genetic stability in a cell in which mismatch repair is blocked via a chemical agent. The chemical is removed from the cell culture and the cell restores its genetic stability.

In another embodiment of the invention, a method is provided for restoring genetic stability in a cell with blocked mismatch repair and a newly selected phenotype. The chemical agent is removed from the cell culture and the cell restores its genetic stability and the new phenotype is stable.

In another embodiment of the invention, a chemical method for blocking MMR in plants is provided. The plant is grown in the presence of a chemical agent. The plant is grown and exhibits an enhanced rate of hypermutation.

In another embodiment of the invention, a method for screening chemical inhibitors of MMR in plants *in vivo* is provided. MMR-sensitive plant expression vectors are engineered. The reporter vectors are introduced into plant hosts. The plant is grown in the presence of a chemical agent. The plant is monitored for altered reporter gene function.

In another embodiment of the invention, a method is provided for introducing a mutation into a gene of interest in a plant. A chemical that blocks mismatch repair is added to a plant. The plant becomes hypermutable as a result of the introduction of the chemical. The plant further comprises a gene of interest. The plant is grown. The plant is tested to determine whether the gene of interest harbors a mutation.

In another embodiment of the invention, a method is provided for producing new phenotypes of a plant. A chemical that blocks mismatch repair is added to a plant. The plant becomes hypermutable as a result of the introduction of the chemical. The plant is grown and tested for the expression of new phenotypes.

In another embodiment of the invention, a method is provided for restoring genetic stability in a plant in which mismatch repair is blocked via a chemical agent. The chemical is removed from the plant culture and the plant restores its genetic stability.

In another embodiment of the invention, a method is provided for restoring genetic stability in a plant with blocked mismatch repair and a newly selected phenotype. The chemical agent is removed from the plant culture and the plant restores its genetic stability and the new phenotype is stable.

These and other embodiments of the invention provide the art with methods that can generate enhanced mutability in microbes, organisms of the protista class, insect cells, mammalian cells, plants, and animals as well as providing cells, plants and animals harboring potentially useful mutations.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** shows diagrams of mismatch repair (MMR) sensitive reporter genes.

**Figure 2** shows a screening method for identifying MMR blocking chemicals.

**Figure 3** shows identification of a small chemical that blocks MMR and genetically alters the pCAR-OF vector *in vivo*.

**Figure 4** shows shifting of endogenous microsatellites in human cells induced by a chemical inhibitor of MMR.

**Figure 5** shows sequence analysis of microsatellites from cells treated with chemical inhibitors of MMR with altered repeats.

**Figure 6** shows generation of host organisms with new phenotypes using a chemical blocker of MMR.

**Figure 7** shows a schematic diagram of MMR-sensitive reporter gene for plants.

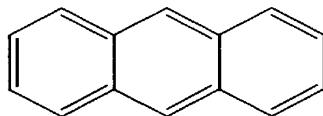
**Figure 8** shows derivatives of lead compounds and thereof that are inhibitors of MMR *in vivo*.

#### DETAILED DESCRIPTION OF THE INVENTION

Various definitions are provided herein. Most words and terms have the meaning that would be attributed to those words by one skilled in the art. Words or terms specifically

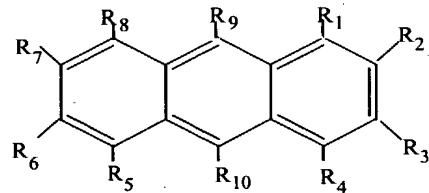
defined herein have the meaning provided in the context of the present invention as a whole and as are typically understood by those skilled in the art. Any conflict between an art-understood definition of a word or term and a definition of the word or term as specifically taught herein shall be resolved in favor of the latter. Headings used herein are for convenience and are not to be construed as limiting.

As used herein the term "anthracene" refers to the compound anthracene. However, when referred to in the general sense, such as "anthracenes," "an anthracene" or "the anthracene," such terms denote any compound that contains the fused triphenyl core structure of anthracene, i.e.,



regardless of extent of substitution.

In certain preferred embodiments of the invention, the anthracene has the formula:



wherein R<sub>1</sub>-R<sub>10</sub> are independently hydrogen, hydroxyl, amino, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, O-alkyl, S-alkyl, N-alkyl, O-alkenyl, S-alkenyl, N-alkenyl, O-alkynyl, S-alkynyl, N-alkynyl, aryl, substituted aryl, aryloxy, substituted aryloxy, heteroaryl, substituted heteroaryl, aralkyloxy, arylalkyl, alkylaryl, alkylaryloxy, arylsulfonyl, alkylsulfonyl, alkoxy carbonyl, aryloxycarbonyl, guanidino, carboxy, an alcohol, an amino acid, sulfonate, alkyl sulfonate, CN, NO<sub>2</sub>, an aldehyde group, an ester, an ether, a crown ether, a ketone, an organosulfur compound, an organometallic group, a carboxylic acid, an organosilicon or a carbohydrate that optionally contains one or more alkylated hydroxyl groups;

wherein said heteroalkyl, heteroaryl, and substituted heteroaryl contain at least one heteroatom that is oxygen, sulfur, a metal atom, phosphorus, silicon or nitrogen; and

wherein said substituents of said substituted alkyl, substituted alkenyl, substituted alkynyl,

substituted aryl, and substituted heteroaryl are halogen, CN, NO<sub>2</sub>, lower alkyl, aryl, heteroaryl, aralkyl, aralkyloxy, guanidino, alkoxycarbonyl, alkoxy, hydroxy, carboxy and amino;

and wherein said amino groups optionally substituted with an acyl group, or 1 to 3 aryl or lower alkyl groups;

or wherein any two of R<sub>1</sub>-R<sub>10</sub> can together form a polyether;

or wherein any two of R<sub>1</sub>-R<sub>10</sub> can, together with the intervening carbon atoms of the anthracene core, form a crown ether.

As used herein, "alkyl" refers to a hydrocarbon containing from 1 to about 20 carbon atoms. Alkyl groups may straight, branched, cyclic, or combinations thereof. Alkyl groups thus include, by way of illustration only, methyl, ethyl, propyl, isopropyl, butyl, isobutyl, cyclopentyl, cyclopentylmethyl, cyclohexyl, cyclohexylmethyl, and the like. Also included within the definition of "alkyl" are fused and/or polycyclic aliphatic cyclic ring systems such as, for example, adamantane. As used herein the term "alkenyl" denotes an alkyl group having at least one carbon-carbon double bond. As used herein the term "alkynyl" denotes an alkyl group having at least one carbon-carbon triple bond.

In some preferred embodiments, the alkyl, alkenyl, alkynyl, aryl, aryloxy, and heteroaryl substituent groups described above may bear one or more further substituent groups; that is, they may be "substituted". In some preferred embodiments these substituent groups can include halogens (for example fluorine, chlorine, bromine and iodine), CN, NO<sub>2</sub>, lower alkyl groups, aryl groups, heteroaryl groups, aralkyl groups, aralkyloxy groups, guanidino, alkoxycarbonyl, alkoxy, hydroxy, carboxy and amino groups. In addition, the alkyl and aryl portions of aralkyloxy, arylalkyl, arylsulfonyl, alkylsulfonyl, alkoxycarbonyl, and aryloxycarbonyl groups also can bear such substituent groups. Thus, by way of example only, substituted alkyl groups include, for example, alkyl groups fluoro-, chloro-, bromo- and iodoalkyl groups, aminoalkyl groups, and hydroxyalkyl groups, such as hydroxymethyl, hydroxyethyl, hydroxypropyl, hydroxybutyl, and the like. In some preferred embodiments such hydroxyalkyl groups contain from 1 to about 20 carbons.

As used herein the term "aryl" means a group having 5 to about 20 carbon atoms and

which contains at least one aromatic ring, such as phenyl, biphenyl and naphthyl. Preferred aryl groups include unsubstituted or substituted phenyl and naphthyl groups. The term "aryloxy" denotes an aryl group that is bound through an oxygen atom, for example a phenoxy group.

In general, the prefix "hetero" denotes the presence of at least one hetero (i.e., non-carbon) atom, which is in some preferred embodiments independently one to three O, N, S, P, Si or metal atoms. Thus, the term "heteroaryl" denotes an aryl group in which one or more ring carbon atom is replaced by such a heteroatom. Preferred heteroaryl groups include pyridyl, pyrimidyl, pyrrolyl, furyl, thienyl, and imidazolyl groups.

The term "aralkyl" (or "arylalkyl") is intended to denote a group having from 6 to 15 carbons, consisting of an alkyl group that bears an aryl group. Examples of aralkyl groups include benzyl, phenethyl, benzhydryl and naphthylmethyl groups.

The term "alkylaryl" (or "alkaryl") is intended to denote a group having from 6 to 15 carbons, consisting of an aryl group that bears an alkyl group. Examples of aralkyl groups include methylphenyl, ethylphenyl and methylnaphthyl groups.

The term "arylsulfonyl" denotes an aryl group attached through a sulfonyl group, for example phenylsulfonyl. The term "alkylsulfonyl" denotes an alkyl group attached through a sulfonyl group, for example methylsulfonyl.

The term "alkoxycarbonyl" denotes a group of formula  $-C(=O)-O-R$  where R is alkyl, alkenyl, or alkynyl, where the alkyl, alkenyl, or alkynyl portions thereof can be optionally substituted as described herein.

The term "aryloxycarbonyl" denotes a group of formula  $-C(=O)-O-R$  where R is aryl, where the aryl portion thereof can be optionally substituted as described herein.

The terms "arylalkyloxy" or "aralkyloxy" are equivalent, and denote a group of formula  $-O-R'-R''$ , where R' is R is alkyl, alkenyl, or alkynyl which can be optionally substituted as described herein, and wherein R'' denotes a aryl or substituted aryl group.

The terms "alkylaryloxy" or "alkaryloxy" are equivalent, and denote a group of formula  $-O-R'-R''$ , where R' is an aryl or substituted aryl group, and R'' is alkyl, alkenyl, or alkynyl which can be optionally substituted as described herein.

As used herein, the term "aldehyde group" denotes a group that bears a moiety of formula  $-C(=O)-H$ . The term "ketone" denotes a moiety containing a group of formula  $-R-C(=O)-R=$ ,

where R and R= are independently alkyl, alkenyl, alkynyl, aryl, heteroaryl, aralkyl, or alkaryl, each of which may be substituted as described herein.

As used herein, the term "ester" denotes a moiety having a group of formula -R-C(=O)-O-R= or -R-O-C(=O)-R= where R and R= are independently alkyl, alkenyl, alkynyl, aryl, heteroaryl, aralkyl, or alkaryl, each of which may be substituted as described herein.

The term "ether" denotes a moiety having a group of formula -R-O-R= or where R and R= are independently alkyl, alkenyl, alkynyl, aryl, heteroaryl, aralkyl, or alkaryl, each of which may be substituted as described herein.

The term "crown ether" has its usual meaning of a cyclic ether containing several oxygen atoms. As used herein the term "organosulfur compound" denotes aliphatic or aromatic sulfur containing compounds, for example thiols and disulfides. The term "organometallic group" denotes an organic molecule containing at least one metal atom.

The term "organosilicon compound" denotes aliphatic or aromatic silicon containing compounds, for example alkyl and aryl silanes.

The term "carboxylic acid" denotes a moiety having a carboxyl group, other than an amino acid.

As used herein, the term "amino acid" denotes a molecule containing both an amino group and a carboxyl group. In some preferred embodiments, the amino acids are  $\alpha$ -,  $\beta$ -,  $\gamma$ - or  $\delta$ -amino acids, including their stereoisomers and racemates. As used herein the term "L-amino acid" denotes an  $\alpha$ -amino acid having the L configuration around the  $\alpha$ -carbon, that is, a carboxylic acid of general formula CH(COOH)(NH<sub>2</sub>)-(side chain), having the L-configuration.

The term "D-amino acid" similarly denotes a carboxylic acid of general formula CH(COOH)(NH<sub>2</sub>)-(side chain), having the D-configuration around the  $\alpha$ -carbon. Side chains of L-amino acids include naturally occurring and non-naturally occurring moieties. Non-naturally occurring (i.e., unnatural) amino acid side chains are moieties that are used in place of naturally occurring amino acid side chains in, for example, amino acid analogs. See, for example, Lehninger, *Biochemistry*, Second Edition, Worth Publishers, Inc, 1975, pages 72-77, incorporated herein by reference. Amino acid substituents may be attached through their carbonyl groups through the oxygen or carbonyl carbon thereof, or through their amino groups, or through functionalities residing on their side chain portions.

As used herein "polynucleotide" refers to a nucleic acid molecule and includes genomic DNA cDNA, RNA, mRNA and the like.

As used herein "antisense oligonucleotide" refers to a nucleic acid molecule that is complementary to at least a portion of a target nucleotide sequence of interest and specifically hybridizes to the target nucleotide sequence under physiological conditions.

As used herein "inhibitor of mismatch repair" refers to an agent that interferes with at least one function of the mismatch repair system of a cell and thereby renders the cell more susceptible to mutation.

As used herein "hypermutable" refers to a state in which a cell *in vitro* or *in vivo* is made more susceptible to mutation through a loss or impairment of the mismatch repair system.

As used herein "agents," "chemicals," and "inhibitors" when used in connection with inhibition of MMR refers to chemicals, oligonucleotides, analogs of natural substrates, and the like that interfere with normal function of MMR.

Methods for developing hypermutable cells and whole organisms have been discovered by taking advantage of the conserved mismatch repair (MMR) process of a host. Dominant negative alleles of MMR genes, when introduced into cells or transgenic animals, increase the rate of spontaneous mutations by reducing the effectiveness of DNA repair and thereby render the cells or animals hypermutable. Hypermutable microbes, protozoans, insects, mammalian cells, plants or whole animals can then be utilized to develop new mutations in a gene of interest. It has been discovered that chemicals that block MMR, and thereby render cells hypermutable, is an efficient way to introduce mutations in cells and genes of interest. In addition to destabilizing the genome of cells exposed to chemicals that inhibit MMR activity may be done transiently, allowing cells to become hypermutable, and removing the chemical exposure after the desired effect (e.g., a mutation in a gene of interest) is achieved. The chemicals that inhibit MMR activity that are suitable for use in the invention include, but are not limited to, anthracene derivatives, nonhydrolyzable ATP analogs, ATPase inhibitors, antisense oligonucleotides that specifically anneal to polynucleotides encoding mismatch repair proteins, DNA polymerase inhibitors, and exonuclease inhibitors. These chemicals can enhance the rate of mutation due to inactivation

of MMR yielding clones or subtypes with altered biochemical properties. Methods for identifying chemical compounds that inhibit MMR *in vivo* are also described herein.

The process of MMR, also called mismatch proofreading, is carried out by a group of protein complexes in cells ranging from bacteria to man (Harfe B.D., and S. Jinks-Robertson (2000) *An. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). An MMR gene is a gene that encodes for one of the proteins of such a mismatch repair complex.

Although not wanting to be bound by any particular theory of mechanism of action, an MMR complex is believed to detect distortions of the DNA helix resulting from non-complementary pairing of nucleotide bases. The non-complementary base on the newer DNA strand is excised, and the excised base is replaced with the appropriate base, which is complementary to the older DNA strand. In this way, cells eliminate many mutations that occur as a result of mistakes in DNA replication.

Dominant negative alleles cause an MMR defective phenotype even in the presence of a wild-type allele in the same cell. An example of a dominant negative allele of an MMR gene is the human gene *hPMS2-134* (SEQ ID NO:25), which carries a truncating mutation at codon 134 (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). The mutation causes the product of this gene to abnormally terminate at the position of the 134th amino acid, resulting in a shortened polypeptide containing the N-terminal 133 amino acids (SEQ ID NO:24). Such a mutation causes an increase in the rate of mutations, which accumulate in cells after DNA replication. Expression of a dominant negative allele of a mismatch repair gene results in impairment of mismatch repair activity, even in the presence of the wild-type allele.

The MMR process has been shown to be blocked by the use of nonhydrolyzable forms of ATP (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson, K.P. *et al.* (2000) *Biochem.* 39:3176-3183). However, it has not been demonstrated that chemicals can block MMR activity in cells. Such chemicals can be identified by screening cells for defective MMR activity. Cells from bacteria, yeast, fungi, insects, plants, animals, and humans can be screened for defective mismatch repair. Genomic DNA, cDNA, or mRNA from any cell can be analyzed for variations from the wild type sequences in cells or organisms grown in the presence of MMR blocking compounds.

Various techniques of screening can be used. The suitability of such screening assays, whether natural or artificial, for use in identifying hypermutable cells, insects, fungi, plants or animals can be evaluated by testing the mismatch repair activity caused by a compound or a mixture of compounds, to determine if it is an MMR inhibitor.

A cell, a microbe, or a whole organism such as an insect, fungus, plant or animal in which a chemical inhibitor of mismatch repair has been treated will become hypermutable. This means that the spontaneous mutation rate of such cells or whole organism is elevated compared to cells or animals without such treatment. The degree of elevation of the spontaneous mutation rate can be at least 2-fold, 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 200-fold, 500-fold, or 1000-fold that of the normal cell or animal. The use of chemical mutagens such as, but limited to, N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), methane sulfonate, dimethyl sulfonate, O6-methyl benzadine, ethyl methanesulfonate (EMS), methylnitrosourea (MNU), ethylnitrosourea (ENU), etc. can be used in MMR defective cells or whole organisms to increase the rates an additional 10 to 100 fold that of the MMR deficiency itself.

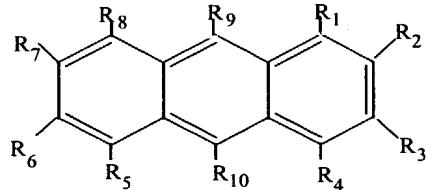
According to one aspect of the invention, a screening assay for identifying chemical inhibitors of MMR is developed and employed. A chemical compound can be in any form or class ranging from but not limited to amino acid, steroid, aromatic, or lipid precursors. The chemical compound can be naturally occurring or made in the laboratory. The screening assay can be natural such as looking for altered endogenous repeats within an host organism's genome (as demonstrated in Figs. 4 and 5), or made in the laboratory using an MMR-sensitive reporter gene as demonstrated in Figs. 1-3).

The chemical compound can be introduced into the cell by supplementing the growth medium, or by intracellular delivery such as but not limited to using microinjection or carrier compounds.

According to another aspect of the invention, a chemical compound from the anthracene class can be exposed to MMR proficient cells or whole organism hosts, the host is grown and screened for subtypes containing genetically altered genes with new biochemical features.

The anthracene compounds that are suitable for use in the invention include, but are

not limited to anthracenes having the formula:



wherein  $R_1-R_{10}$  are independently hydrogen, hydroxyl, amino, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, O-alkyl, S-alkyl, N-alkyl, O-alkenyl, S-alkenyl, N-alkenyl, O-alkynyl, S-alkynyl, N-alkynyl, aryl, substituted aryl, aryloxy, substituted aryloxy, heteroaryl, substituted heteroaryl, aralkyloxy, arylalkyl, alkylaryl, alkylaryloxy, arylsulfonyl, alkylsulfonyl, alkoxy carbonyl, aryloxycarbonyl, guanidino, carboxy, an alcohol, an amino acid, sulfonate, alkyl sulfonate, CN, NO<sub>2</sub>, an aldehyde group, an ester, an ether, a crown ether, a ketone, an organosulfur compound, an organometallic group, a carboxylic acid, an organosilicon or a carbohydrate that optionally contains one or more alkylated hydroxyl groups;

wherein said heteroalkyl, heteroaryl, and substituted heteroaryl contain at least one heteroatom that is oxygen, sulfur, a metal atom, phosphorus, silicon or nitrogen; and

wherein said substituents of said substituted alkyl, substituted alkenyl, substituted alkynyl,

substituted aryl, and substituted heteroaryl are halogen, CN, NO<sub>2</sub>, lower alkyl, aryl, heteroaryl, aralkyl, aralkyloxy, guanidino, alkoxy carbonyl, alkoxy, hydroxy, carboxy and amino;

and wherein said amino groups optionally substituted with an acyl group, or 1 to 3 aryl or lower alkyl groups;

or wherein any two of  $R_1-R_{10}$  can together form a polyether; or wherein any two of  $R_1-R_{10}$  can, together with the intervening carbon atoms of the anthracene core, form a crown ether.

The method of the invention also encompasses inhibiting MMR with an anthracene of the above formula wherein  $R_5$  and  $R_6$  are hydrogen, and the remaining substituents are as described above.

The some embodiments, in the anthracene compound  $R_1-R_{10}$  are independently hydrogen, hydroxyl, alkyl, aryl, arylalkyl, or hydroxyalkyl. In other embodiments,  $R_1-R_{10}$  are independently

hydrogen, hydroxyl, methyl, ethyl, propyl, isopropyl, butyl, isobutyl, phenyl, toyl, hydroxymethyl, hydroxypropyl, or hydroxybutyl.

In specific embodiments of the invention the anthracenes include, but are not limited to 1,2-dimethylanthracene, 9,10-dimethyl anthracene, 7,8-dimethylanthracene, 9,10-diphenylanthracene, 9,10-dihydroxymethylanthracene, 9-hydroxymethyl-10-methylanthracene, dimethylanthracene-1,2-diol, 9-hydroxymethyl-10-methylanthracene-1,2-diol, 9-hydroxymethyl-10-methylanthracene-3,4-diol, 9, 10-di-m-tolyanthracene, and the like.

The chiral position of the side chains of the anthracenes is not particularly limited and may be any chiral position and any chiral analog. The anthracenes may also comprise a stereoisomeric forms of the anthracenes and includes any isomeric analog.

Examples of hosts are but not limited to cells or whole organisms from human, primate, mammal, rodent, plant, fish, reptiles, amphibians, insects, fungi, yeast or microbes of prokaryotic origin.

Yet another aspect of the invention is the use of ATP analogs capable of blocking ATPase activity required for MMR. MMR reporter cells are screened with ATP compound libraries to identify those compounds capable of blocking MMR *in vivo*. Examples of ATP analogs that are useful in blocking MMR activity include, but are not limited to, nonhydrolyzable forms of ATP such as AMP-PNP and ATP[gamma]S block the MMR activity (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson K.P. *et al.* (2000) *Biochem.* 39:3176-3183).

Yet another aspect of the invention is the use of nuclease inhibitors that are able to block the exonuclease activity of the MMR biochemical pathway. MMR reporter cells are screened with nuclease inhibitor compound libraries to identify compounds capable of blocking MMR *in vivo*. Examples of nuclease inhibitors that are useful in blocking MMR activity include, but are not limited to analogs of N-Ethylmaleimide, an endonuclease inhibitor (Huang, Y.C., et.al. (1995) *Arch. Biochem. Biophys.* 316:485), heterodimeric adenine-chain-acridine compounds, exonuclease III inhibitors (Belmont P, et.al., *Bioorg Med Chem Lett* (2000) 10:293-295), as well as antibiotic compounds such as Heliquinomycin, which have helicase inhibitory activity (Chino, M, et.al. *J. Antibiot. (Tokyo)* (1998) 51:480-486).

Another aspect of the invention is the use of DNA polymerase inhibitors that are able to block the polymerization required for mismatch-mediated repair. MMR reporter cells are screened with DNA polymerase inhibitor compound libraries to identify those compounds capable of blocking MMR *in vivo*. Examples of DNA polymerase inhibitors that are useful in blocking MMR activity include, but are not limited to, analogs of actinomycin D (Martin, S.J., et.al. (1990) *J. Immunol.* 145:1859), Aphidicolin (Kuwakado, K. et.al. (1993) *Biochem. Pharmacol.* 46:1909) 1-(2'-Deoxy-2'-fluoro-beta-L-arabinofuranosyl)-5-methyluracil (L-FMAU) (Kukhanova M, et.al., *Biochem Pharmacol* (1998) 55:1181-1187), and 2',3'-dideoxyribonucleoside 5'-triphosphates (ddNTPs) (Ono, K., et.al., *Biomed Pharmacother* (1984) 38:382-389).

In yet another aspect of the invention, antisense oligonucleotides are administered to cells to disrupt at least one function of the mismatch repair process. The antisense polynucleotides hybridize to MMR polynucleotides. Both full-length and antisense polynucleotide fragments are suitable for use. "Antisense polynucleotide fragments" of the invention include, but are not limited to polynucleotides that specifically hybridize to an MMR encoding RNA (as determined by sequence comparison of nucleotides encoding the MMR to nucleotides encoding other known molecules). Identification of sequences that are substantially unique to MMR-encoding polynucleotides can be ascertained by analysis of any publicly available sequence database and/or with any commercially available sequence comparison programs. Antisense molecules may be generated by any means including, but not limited to chemical synthesis, expression in an *in vitro* transcription reaction, through expression in a transformed cell comprising a vector that may be transcribed to produce antisense molecules, through restriction digestion and isolation, through the polymerase chain reaction, and the like.

Antisense oligonucleotides, or fragments thereof may include the nucleotide sequences set forth in SEQ ID NOS:15, 17, 19, 21, 23, 25, 27, and 29 or sequences complementary or homologous thereto, for example. Those of skill in the art recognize that the invention may be predicted using any MMR gene. Specifically, antisense nucleic acid molecules comprise a sequence complementary to at least about 10, 15, 25, 50, 100, 250 or 500 nucleotides or an entire MMR encoding sequence. Preferably, the antisense

oligonucleotides comprise a sequence complementary to about 15 consecutive nucleotides of the coding strand of the MMR encoding sequence.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MMR protein. The coding strand may also include regulatory regions of the MMR sequence. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of human PMS2 corresponds to the coding region SEQ ID NO:17). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding an MMR protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions (UTR)).

Preferably, antisense oligonucleotides are directed to regulatory regions of a nucleotide sequence encoding an MMR protein, or mRNA corresponding thereto, including, but not limited to, the initiation codon, TATA box, enhancer sequences, and the like. Given the coding strand sequences provided herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an MMR mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of an MMR mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of an MMR mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length.

Screening is any process whereby a chemical compound is exposed to a cell or whole organism. The process of screening can be carried out using but not limited to a whole animal, plant, insect, microbe, or by using a suspension of one or more isolated cells in culture. The cell can be any type of eukaryotic or prokaryotic cell, including, for example, cells isolated from humans or other primates, mammals or other vertebrates, invertebrates, and single celled organisms such as protozoa, yeast, or bacteria.

In general, screening will be carried out using a suspension of cells, or a single cell, but other methods can also be applied as long as a sufficient fraction of the treated cells or tissue is exposed so that isolated cells can be grown and utilized. Techniques for chemical screening are well known to those in the art. Available techniques for screening include cell-based assays, molecular assays, and whole organism-based assays. Compounds can be added to the screening assays of the invention in order to identify those agents that are capable of blocking MMR in cells.

The screening assays of the invention provide a system wherein a cell, cells or a whole organism is contacted with a candidate compound and then tested to determine whether mismatch repair has been adversely affected. The method in which MMR is analyzed may be any known method, including, but not limited to analysis of the molecular sequence of the MMR gene, and analyzing endogenous repeats in the subject's genome. Further, the invention provides a convenient assay to analyze the effects of candidate agents on reporter genes transfected into cells.

MMR-inhibitors identified by the methods of the invention can be used to generate new mutations in one or more gene(s) of interest. A gene of interest can be any gene naturally possessed by a cell line, microbe or whole organism. An advantage of using chemicals rather than recombinant technologies to block MMR are that the process is faster; there is no need to produce stable clones with a knocked out MMR gene or a clone expressing a dominant negative MMR gene allele. Another advantage is that host organisms need not be screened for integrated knock out targeting vectors or stable expression of a dominant negative MMR gene allele. Finally, once a cell, plant or animal has been exposed to the MMR-blocking compound and a new output trait is generated, the MMR process can be restored by removal of compound. Mutations can be detected by analyzing the genotype of the cell, or whole organism, for example, by examining the sequence of genomic DNA, cDNA, messenger RNA, or amino acids associated with the gene of interest. Mutations can also be detected by screening for new output traits such as hypoxanthine-guanine phosphoribosyltransferase (HPRT) revertants. A mutant polypeptide can be detected by identifying alterations in electrophoretic mobility, spectroscopic properties, or other physical or structural characteristics of a protein encoded by a mutant gene. One can also screen for

altered function of the protein *in situ*, in isolated form, or in model systems. One can screen for alteration of any property of the cell, plant or animal associated with the function of the gene of interest.

Several advantages exist in generating genetic mutations by blocking MMR *in vivo* in contrast to general DNA damaging agents such as MNNG, MNU and EMS. Cells with MMR deficiency have a wide range of mutations dispersed throughout their entire genome in contrast to DNA damaging agents such as MNNG, MNU, EMS and ionizing radiation. Another advantage is that mutant cells that arise from MMR deficiency are diploid in nature and do not lose large segments of chromosomes as is the case of DNA damaging agents such as EMS, MNU, and ionizing radiation (Honma, M. *et al.* (1997) *Mutat. Res.* 374:89-98). This unique feature allows for subtle changes throughout a host's genome that leads to subtle genetic changes yielding genetically stable hosts with commercially important output traits.

The invention also encompasses blocking MMR *in vivo* and *in vitro* and further exposing the cells or organisms to a chemical mutagen in order to increase the incidence of genetic mutation.

The invention also encompasses withdrawing exposure to inhibitors of mismatch repair once a desired mutant genotype or phenotype is generated such that the mutations are thereafter maintained in a stable genome.

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific examples, which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

## EXAMPLES

### **EXAMPLE 1: Generation of a cell-based screening assay to identify chemicals capable of inactivating mismatch repair *in vivo*.**

A hallmark of MMR deficiency is the generation of unstable microsatellite repeats in the genome of host cells (Peinado, M.A. *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Strand, M. *et al.* (1993) *Nature* 365:274-276; Parsons, R. *et al.* (1993) *Cell* 75:1227-1236). This phenotype is referred to as microsatellite instability (MI) (Harfe, B.D.

and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960; Peinado, M.A. et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Perucho, M. (1996) *Biol. Chem.* 377:675-684; Hoang, J.M. et al. (1997) *Cancer Res.* 57:300-303; Strand, M. et al. (1993) *Nature* 365:274-276). MI consists of deletions and/or insertions within repetitive mono-, di- and/or tri nucleotide repetitive sequences throughout the entire genome of a host cell. Extensive genetic analysis of eukaryotic cells have found that the only biochemical defect that is capable of producing MI is defective MMR (Harfe, B.D. and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960; Peinado, M.A. et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10065-10069; Perucho, M. (1996) *Biol. Chem.* 377:675-684; Hoang, J.M. et al. (1997) *Cancer Res.* 57:300-303; Strand, M. et al. (1993) *Nature* 365:274-276). In light of this unique feature that defective MMR has on promoting microsatellite instability, endogenous MI is now used as a biochemical marker to survey for lack of MMR activity within host cells (Hoang, J.M. et al. (1997) *Cancer Res.* 57:300-303).

A method used to detect MMR deficiency in eukaryotic cells is to employ a reporter gene that has a polynucleotide repeat inserted within the coding region that disrupts its reading frame due to a frame shift. In the case where MMR is defective, the reporter gene will acquire random mutations (*i.e.*, insertions and/or deletions) within the polynucleotide repeat yielding clones that contain a reporter with an open reading frame. This reporter gene can be of any biochemical pathway such as but not limited to  $\beta$ -glucuronidase,  $\beta$ -galactosidase, neomycin resistant gene, hygromycin resistance gene, green fluorescent protein, and the like. A schematic diagram of MMR-sensitive reporters are shown in Fig. 1, where the polynucleotide repeat can consist of mono-, di-, tri- or tetra-nucleotides. We have employed the use of a  $\beta$ -galactosidase MMR-sensitive reporter gene to measure for MMR activity in H36 cells, which are a murine hybridoma cell line. The reporter construct used is called pCAR-OF, which contains a hygromycin resistance (HYG) gene plus a  $\beta$ -galactosidase gene with a 29 bp out-of-frame poly-CA tract inserted at the 5' end of its coding region. The pCAR-OF reporter cannot generate  $\beta$ -galactosidase activity unless a frame-restoring mutation (*i.e.*, insertion or deletion) arises following transfection. This line has been shown to be sensitive to inactivated MMR where using a dominant negative MMR gene allele has found

this condition to result in the production of  $\beta$ -galactosidase (unpublished data). An example of these data using the dominant negative PMS134 allele is shown in Table 1. Briefly, H36 cells were each transfected with an expression vector containing the PMS134 allele (referred to as HB134) or empty vector and the pCAR-OF vector in duplicate reactions using the protocol below. The PMS134 gene is cloned into the pEF expression vector, which contains the elongation factor promoter upstream of the cloning site followed by a mammalian polyadenylation signal. This vector also contains the NEOr gene that allows for selection of cells in G418 to identify those retaining this plasmid. Briefly, cells were transfected with 1  $\mu$ g of the PMS134 or empty vector using polyliposomes following the manufacturer's protocol (Life Technologies). Cells were then selected in 0.5 mg/ml of G418 for 10 days and G418 resistant cells were pooled together to analyze for gene expression. PMS134 positive cells, which were determined by RT-PCR and western blot (not shown) were expanded and transfected with the pCAR-OF reporter gene that contains a hygromycin (HYG) resistance gene as reporter using the protocol described above. Cells were selected in 0.5 mg/ml G418 and 0.5mg/ml HYG to select for cells retaining both the MMR effector and the pCAR-OF reporter plasmids. All cultures transfected with the pCAR vector resulted in a similar number of HYG/G418 resistant cells. Cultures were then expanded and tested for  $\beta$ -galactosidase activity *in situ* as well as by biochemical analysis of cell extracts. For *in situ* analysis, 100,000 cells were harvested and fixed in 1% gluteraldehyde, washed in phosphate buffered saline solution and incubated in 1 ml of X-gal substrate solution [0.15 M NaCl, 1 mM MgCl<sub>2</sub>, 3.3 mM K<sub>4</sub>Fe(CN)<sub>6</sub>, 3.3 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 0.2% X-Gal ] in 24 well plates for 2 hours at 37°C. Reactions were stopped in 500 mM sodium bicarbonate solution and transferred to microscope slides for analysis. Three fields of 200 cells each were counted for blue ( $\beta$ -galactosidase positive cells) or white ( $\beta$ -galactosidase negative cells) to assess for MMR inactivation. Table 1 shows the results from these studies. While no  $\beta$ -galactosidase positive cells were observed in H36 empty vector cells and 10% of the cells per field were  $\beta$ -galactosidase positive in HB134 cultures.

**Table 1.**  $\beta$ -galactosidase expression of H36 empty vector and HB134 cells transfected with pCAR-OF reporter vectors. Cells were transfected with the pCAR-OF reporter plasmid.

Transfected cells were selected in HYG and G418, expanded and stained with X-gal solution to measure for  $\beta$ -galactosidase activity (blue colored cells). 3 fields of 200 cells each were analyzed by microscopy. The results below represent the mean +/- standard deviation of these experiments.

Table 1.

CELL LINE	# BLUE CELLS
H36 empty vector	0 +/- 0
HB134	20 +/- 3

Cultures can be further analyzed by biochemical assays using cell extracts to measure  $\beta$ -galactosidase activity as previously described (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641).

The data described in Table 1 show that by inhibiting the MMR activity of an MMR proficient cell host can result in MI and the altering of microsatellites in the pCAR-OF vector results in cells that produce functional  $\beta$ -galactosidase enzyme. The use of the H36pCAR-OF cell line can now be used to screen for chemicals that are able to block MMR of the H36 cell line.

#### **EXAMPLE 2: Screening assays for identifying chemical blockers of MMR.**

A method for screening chemical libraries is provided in this example using the H36pCAR-OF cell line described in Example 1. This cell line is a hardy, stable line that can be formatted into 96-well microtiter plates for automated screening for chemicals that specifically block MMR. An overview of the screening process is given in Figure 2, however, the process is not limited to the specifications within this example. Briefly, 10,000 cells in a total volume of 0.1ml of growth medium (RPMI1640 plus 10% fetal bovine serum) are added to 96-well microtiter plates containing any variety of chemical compounds. Cells are grown for 14-17 days at 37°C in 5%CO<sub>2</sub>. Cells are then lysed in the growth medium with 50uls of lysis buffer containing 0.1 M Tris buffer (pH 8.0), 0.1% Triton X-100, 45 mM 2-mercaptoethanol, 1mM MgCl<sub>2</sub>, 0.1 M NaPO<sub>4</sub> and 0.6 mg/ml Chlorophenol-red-  $\beta$ -

D-galactopyranoside (CPRG, Roche). Reactions are incubated for 1 hour, terminated by the addition of 50  $\mu$ ls of 0.5 M Na<sub>2</sub>CO<sub>3</sub>, and analyzed by spectrophotometry at 576 nm.

Experimental wells are compared to untreated or vehicle treated wells to identify those with increased  $\beta$ -galactosidase activity. Compounds producing MMR blocking activity are then further analyzed using different cell lines containing the pCAR-OF plasmid to measure the ability to block MMR as determined by MI in MMR proficient hosts by analyzing endogenous microsatellites for instability using assays described below.

#### **EXAMPLE 3: Defining MMR blocking chemicals.**

The identification of chemical inhibitors of MMR can be difficult in determining those that are standard mutagens from those that induce genomic instability via the blockade of MMR. This Example teaches of a method for determining blockers of MMR from more general mutagens. Once a compound has been identified in the assay described above, one can determine if the compound is a general mutagen or a specific MMR blocker by monitoring mutation rates in MMR proficient cells and a controlled subclone that is MMR defective. One feature of MMR deficiency is the increased resistance to toxicity of DNA alkylating agents that allows for enhanced rates of mutations upon mutagen exposure (Liu, L., et.al. *Cancer Res* (1996) 56:5375-5379). This unique feature allows for the use of a MMR proficient cell and a controlled line to measure for enhanced activity of a chemical compound to induce mutations in MMR proficient vs MMR deficient lines. If the compound is a true inhibitor of MMR then genetic mutations should occur in MMR proficient cells while no "enhanced" mutation rate will be found in already MMR defective cells. Using these criteria chemicals such as ICR191, which induces frameshift mutations in mammalian cells would not be considered a MMR blocking compound because of its ability to produce enhanced mutation rates in already MMR defective cell lines (Chen, W.D., et.al. *J Natl Cancer Inst.* (2000) 92:480-485). These screening lines include the but are not limited those in which a dominant negative MMR gene has been introduced such as that described in EXAMPLE 1 or those in which naturally MMR deficient cells such as HCT116 has been cured by introduction of a complementing MMR gene as described (Chen, W.D., et.al. *J Natl Cancer Inst.* (2000) 92:480-485).

**EXAMPLE 4: Identification of chemical inhibitors of MMR *in vivo*.**

MMR is a conserved post replicative DNA repair mechanism that repairs point mutations and insertion/deletions in repetitive sequences after cell division. The MMR requires an ATPase activity for initiation complex recognition and DNA translocation. *In vitro* assays have shown that the use of nonhydrolyzable forms of ATP such as AMP-PNP and ATP[gamma]S block the MMR activity (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson K.P. *et al.* (2000) *Biochem.* 39:3176-3183).

The use of chemicals to inhibit endogenous MMR *in vivo* has not been distinguished in the public domain. In an attempt to identify chemicals that can inhibit MMR *in vivo*, we used our H36pCAR-OF screening assay to screen for chemicals that are able to cause microsatellite instability and restoration of  $\beta$ -galactosidase activity from the pCAR-OF vector, an effect that can only be caused due to MMR deficiency. In our screening assays we used a variety of classes of compounds ranging from steroids such as ponasterone to potent alkylating agents such as EMS, to kinase and other enzyme inhibitors. Screens identified one class of chemicals that were capable of generating  $\beta$ -galactosidase positive cells. These molecules were derived from the anthracene class. An example of one such anthracene derivative for the purposes of this application is a molecule called 9,10-dimethylanthracene, referred to from here on as DMA. Fig. 3 shows the effect of DMA in shifting the pCAR-OF reporter plasmid. In contrast, general DNA alkylating agents such as EMS or MNNG did not result in MI and/or the shifting of the polynucleotide tract in the pCAR-OF reporter.

The most likely explanation for the differences in  $\beta$ -galactosidase activity was that the DMA compound disturbed MMR activity, resulting in a higher frequency of mutation within the pCAR-OF reporter and re-establishing the ORF. To directly test the hypothesis that MMR was altered, we employ a biochemical assay for MMR with the individual clones as described by Nicolaides *et al.*, 1997 (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). Nuclear extracts are prepared from the clones and incubated with heteroduplex substrates containing either a /CA\ insertion-deletion or a G/T mismatch under conditions described previously. The /CA\ and G/T heteroduplexes are used to test repair from the 3'

and 5' directions, respectively as described (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641).

### **Biochemical assays for mismatch repair.**

#### **Enzymatic Repair Assays:**

MMR activity in nuclear extracts is performed as described, using 24 fmol of substrate (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). Complementation assays are done by adding ~ 100 ng of purified MutLa or MutSa components to 100 µg of nuclear extract, adjusting the final KCl concentration to 100 mM (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). The substrates used in these experiments contain a strand break 181 nucleotides 5' or 125 nucleotides 3' to the mismatch.

#### **Biochemical Activity Assays:**

To demonstrate the direct effect to small molecules on MMR proteins, molecular assays such as mismatch binding and MMR complex formation are performed in the presence or absence of drug. Briefly, MMR gene cDNAs are PCR amplified using primers encompassing the entire coding regions of the known MMR proteins MSH2 (SEQ ID NO:20), GTBP (SEQ ID NO:26), MLH1 (SEQ ID NO:22), human PMS2 (SEQ ID NO:16), mouse PMS2 (SEQ ID NO:14), PMS1 (SEQ ID NO:18), and MHS3 (SEQ ID NO:28) from any species with a sense primer containing a T7 promoter and a Kozak translation signal as previously described (Nicolaides, N.C. *et al.* (1998) *Mol. Cell. Biol.* 18:1635-1641). The coding regions of known MMR proteins include the sequences shown in Table 3 for mouse *PMS2* (SEQ ID NO:15), human *PMS2* (SEQ ID NO:17), human *PMS1* (SEQ ID NO:19), human *MSH2* (SEQ ID NO:21), human *MLH1* (SEQ ID NO:23), and human *MSH3* (SEQ ID NO:29). Products are transcribed and translated using the TNT system (Promega). An example of PCR primers and *in vitro* transcription-translation reactions are listed below.

**In vitro transcription-translation:**

Linear DNA fragments containing *hPMS2* (SEQ ID NO:17) and *hMLH1* (SEQ ID NO:23) cDNA sequences were prepared by PCR, incorporating sequences for *in vitro* transcription and translation in the sense primer. A full-length *hMLH1* fragment was prepared using the sense primer 5'-ggatcctaatacgactcactatagggagaccacatgtcggtggcagg-3' (SEQ ID NO:1)(codons 1-6) and the antisense primer 5'-taagtcctaagtgcaccaac-3' (SEQ ID NO:2)(located in the 3' untranslated region, nt 2411-2433), using a wild-type *hMLH1* cDNA clone as template. A full-length *hPMS2* fragment was prepared with the sense primer 5'-ggatcctaatacgactcactatagggagaccacatggaaacaattgcctgcgg-3' (SEQ ID NO:3)(codons 1-6) and the antisense primer 5'-aggtagtgaagactctgtc-3' (SEQ ID NO:4)(located in 3' untranslated region, nt 2670-2690) using a cloned *hPMS2* cDNA as template. These fragments were used to produce proteins via the coupled transcription-translation system (Promega). The reactions were supplemented with <sup>35</sup>S-labelled methionine or unlabelled methionine. Lower molecular weight bands are presumed to be degradation products and/or polypeptides translated from alternative internal methionines.

To study the effects of MMR inhibitors, assays are used to measure the formation of MLH1 and PMS2 with or without compound using polypeptides produced in the TNT System (Promega) followed by immunoprecipitation (IP). To facilitate the IP, tags may be placed at the C-terminus of the PMS2 protein to use for antibody binding or antibodies directed to the MMR protein itself can be used for IP.

**Immunoprecipitations:**

Immunoprecipitations are performed on *in vitro* translated proteins by mixing the translation reactions with 1 µg of the MLH1 specific monoclonal antibody (mAB) MLH14 (Oncogene Science, Inc.), a polyclonal antibody generated to codons 2-20 of hPMS2 described above, or a polyclonal antibody generated to codons 843-862 of hPMS2 (Santa Cruz Biotechnology, Inc.) in 400 µl of EBC buffer (50 mM Tris, pH 7.5, 0.1 M NaCl, 0.5% NP40). After incubation for 1 hr at 4°C, protein A sepharose (Sigma) is added to a final concentration of 10% and reactions are incubated at 4°C for 1 hour. Proteins bound to protein

A are washed five times in EBC and separated by electrophoresis on 4-20% Tris-glycine gels, which are then dried and autoradiographed.

Compounds that block heterodimerization of mutS or mutL proteins can now be identified using this assay.

**EXAMPLE 5: Use of chemical MMR inhibitors yields microsatellite instability in human cells**

In order to demonstrate the global ability of a chemical inhibitor of MMR in host cells and organisms, we treated human HEK293 cells (referred to as 293 cells) with DMA and measured for microsatellite instability of endogenous loci using the BAT26 diagnostic marker (Hoang J.M. *et al.* (1997) *Cancer Res.* 57:300-303). Briefly,  $10^5$  cells were grown in control medium or 250  $\mu$ M DMA, a concentration that is found to be non-toxic, for 14 to 17 days. Cells are then harvested and genomic DNA isolated using the salting out method (Nicolaides, N.C. *et al.* (1991) *Mol. Cell. Biol.* 11:6166-6176.).

Various amounts of test DNAs from HCT116 (a human colon epithelial cell line) and 293 were first used to determine the sensitivity of our microsatellite test. The BAT26 alleles are known to be heterogeneous between these two cell lines and the products migrate at different molecular weights (Nicolaides personal observation). DNAs were diluted by limiting dilution to determine the level of sensitivity of the assay. DNAs were PCR amplified using the BAT26F: 5'-tgactactttgacttcagcc-3' (SEQ ID NO:43) and the BAT26R: 5'-aaccattcaacattttaaccc-3' (SEQ ID NO:44) primers in buffers as described (Nicolaides, N.C. *et al.* (1995) *Genomics* 30:195-206). Briefly 1 pg to 100 ngs of DNA were amplified using the following conditions: 94°C for 30 sec, 58°C for 30 sec, 72°C for 30 sec for 30 cycles. PCR reactions were electrophoresed on 12% polyacrylamide TBE gels (Novex) or 4% agarose gels and stained with ethidium bromide. These studies found that 0.1 ng of genomic DNA was the limit of detection using our conditions.

To measure for microsatellite stability in 293 cells grown with or without DMA, 0.1 ngs of DNA from DMA-treated or control 293 cells were amplified using the reaction conditions above. Forty individual reactions were carried out for each sample to measure for minor alleles. Fig. 4A shows a typical result from these studies whereby BAT26 alleles were

amplified from DMA-treated and untreated cells and analyzed on 12% PAGE gels (Novex). Alleles from DMA-treated cells showed the presence of an altered allele (asterisk) that migrated differently from the wild type allele. No altered alleles were found in the MMR-proficient control cells as expected since MI only occurs in MMR defective cell hosts. To confirm these data, PCRs were repeated using isolated BAT26 products. Primers and conditions were the same as described above except that reactions were amplified for 20 cycles. PCR products were gel-purified and cloned into T-tailed vectors (InVitrogen) as suggested by the manufacturer. Recombinant clones from DMA-treated and control cells were screened by PCR again using the BAT26 primers. Fifty bacterial colonies were analyzed for BAT26 structure by directly adding an aliquot of live bacteria to the PCR mix. PCR reactions were carried out as described above, and products were electrophoresed on 4% agarose gels and stained with ethidium bromide. As shown in Figure 4B, microsatellites from DMA-treated cells had alterations (asterisks) that made the marker length larger or smaller than the wild type allele found in control cells.

To confirm that these differences in molecular weight were due to shifts within the polynucleotide repeat, a hallmark of defective MMR, five clones from each sample were sequenced using an ABI automated sequencer with an M13-R primer located in the T-tail vector backbone. Sequence analysis revealed that the control cell clone used in our studies was homozygous for the BAT26 allele with a 26nt polyA repeat. Cells treated with DMA found multiple alleles ranging from the wild-type with 26 polyA repeat to shorter alleles (24 polyA repeat) and larger alleles (28 polyA repeat) (Fig. 5).

These data corroborate the H36pCAR data in Example 1 and Fig. 3 and demonstrates the ability to block MMR with a chemical in a range of hosts.

**Example 6: Chemical inhibitors of MMR generate DNA hypermutability in Plants and new phenotypes.**

To determine if chemical inhibitors of MMR work across a diverse array of organisms, we explored the activity of DMA on *Arabidopsis thaliana* (AT), a member of the mustard plant family, as a plant model system to study the effects of DMA on generating MMR deficiency, genome alterations, and new output traits.

Briefly, AT seeds were sterilized with straight commercial bleach and 100 seeds were plated in 100mm Murashige and Skoog (MS) phytagar (Life Technology) plates with increasing amounts of DMA (ranging from 100 $\mu$ M to 50mM). A similar amount of seeds were plated on MS phytagar only or in MS phytagar with increasing amounts of EMS (100 $\mu$ M to 50mM), a mutagen commonly used to mutate AT seeds (McCallum, C.M. et al. (2000) *Nat. Biotechnol.* 18:455-457). Plates were grown in a temperature-controlled, fluorescent-lighted humidifier (Percival Growth Chamber) for 10 days. After 10 days, seeds were counted to determine toxicity levels for each compound. Table 2 shows the number of healthy cells/treatment as determined by root formation and shoot formation. Plantlets that were identical to untreated seeds were scored healthy. Seeds with stunted root or shoot formation were scored intermediate (inter). Non-germinated seeds were scored dead.

**Table 2: Toxicity curve of DMA and EMS on *Arabidopsis* (per 100 cells)**

	0	0.1	0.5	1.0	2.5	5.0	10	12.5	25	50
<b>DMA</b>										
Healthy	100	94	99	99	80	85	65	0	0	0
Inter	0	0	0	0	20	15	32	85	100	0
Dead	0	0	0	0	0	0	0	0	0	100
<b>EMS</b>										
Healthy	99	100	45	25	0	0	0	0	0	0
Inter	0	0	54	75	0	0	0	0	0	0
Dead	0	0	0	0	100	100	100	100	100	87

The data in Table 2 show that DMA toxicity occurs at 10mM of continuous culture, while toxicity occurs at 250  $\mu$ M for EMS. Next, 50 seeds were plated in two 150mm dishes containing 2mM DMA, 250  $\mu$ M EMS or no drug. Seeds were grown for 10 days and then 10 plants from each plate were transferred to soil. All plants appeared to be similar in color and height. Plants were grown at room temperature with daily cycles of 18 hr light and 6 hr dark.

After 45 days seeds are harvested from siliques and stored in a desiccator at 4°C for 72 hours. Seeds are then sterilized and 100 seeds from each plant is sown directly into water-saturated soil and grown at room temperature with daily cycles of 18 hr light and 6 hr dark. At day 10 phenotypically distinct plants were found in 7 out of 118 DMA treated while no phenotypic difference was observed in 150 EMS-treated or 150 control plants. These 7 altered plants were light green in color and appeared to grow slower. Fig. 6 shows a typical difference between the DMA altered plant and controls. DMA-exposed plants produced offspring that were yellow in appearance in contrast to dark green, which is always found in wild-type plants. In addition, the yellow plants were also shorter. After 30 days, most wild-type plants produced flowers and siliques, while the 7 mutants just began flowering. After 45 days, control plants were harvested while mutant plants were harvested 10 to 15 days later. No such effects were observed in 150 plantlets from EMS treated plants.

The effect of DMA on MMR was confirmed by monitoring the structure of endogenous polynucleotide repeat markers within the plant genome. DNA was extracted using the DNazol method following the manufacturer's protocol (Life Technology). Briefly, two leaves were harvested from DMA, EMS or untreated plants and DNA was extracted. DNAs were quantified by optical density using a BioRad Spectrophotometer. In *Arabidopsis*, a series of poly-A (A)<sub>n</sub>, (CA)<sub>n</sub> and (GA)<sub>n</sub> markers were found as a result of EMBL and GenBank database searches of DNA sequence data generated as a result of the *Arabidopsis* genome-sequencing project. Two markers that are naturally occurring, ATHACS and Nga128 are used to monitor microsatellite stability using primers described (Bell, C.J. and J.R. Ecker (1994) *Genomics* 19:137-144). ATHACS has a stretch of thirty-six adenine repeats (A)<sub>36</sub> whereas Nga128 is characterized by a di-nucleotide AG repeat that is repeated nineteen times (AG)<sub>19</sub>, while the Nga280 marker contains a polyAG repeat marker with 15 dinucleotides. DMA-mediated alterations of these markers are measured by a PCR assay.

Briefly, the genomic DNA is amplified with specific primers in PCR reaction buffers described above using 1-10ng plant genomic DNA. Primers for each marker are listed below:

nga280:

nga280-F: 5'-CTGATCTCACGGACAATAGTGC-3' (SEQ ID NO:5)

nga280-R: 5'-GGCTCCATAAAAAGTGCACC-3' (SEQ ID NO:6)

nga128:

nga128-F: 5'-GGTCTGTTGATGTCGTAAGTCG-3' (SEQ ID NO:7)  
nga128-R: 5'-ATCTGAAACCTTAGGGAGGG-3' (SEQ ID NO:8)

ATHACS:

ATHACS-F: 5'-AGAAGTTAGACAGGTAC-3' (SEQ ID NO:9)  
ATHACS-R: 5'-AAATGTGCAATTGCCTTC-3' (SEQ ID NO:10)

Cycling conditions are 94°C for 15 seconds, 55°C for 15 seconds and 72°C for 30 seconds, conditions that have been demonstrated to efficiently amplify these two markers (personal observation, Morphotek). PCR products are analyzed on 3.5% metaphor agarose gel in Tris-Acetate-EDTA buffer following staining with ethidium bromide.

Another method used to demonstrate that biochemical activity of a plant host's MMR is through the use of reporter genes disrupted by a polynucleotide repeat, similar to that described in Example 1 and Fig. 1. Due to the high endogenous  $\beta$ -galactosidase background, we engineered a plant compatible MMR-sensitive reporter gene consisting of the  $\beta$ -glucoronidase (GUS) gene with a mononucleotide repeat that was inserted just downstream of the initiation codon. Two reporter constructs were generated. pGUS-OF, contained a 20 base adenine repeat inserted just downstream of the initiating methionine that resulted in a frameshift, therefore producing a nonfunctional enzyme. The second, pGUS-IF, contained a 19 base adenine repeat that retained an open reading frame and served as a control for  $\beta$ -glucuronidase activity. Both constructs were generated by PCR using the pBI-121 vector (Life Technologies) as template. The antisense primer was directed to the 3' end of the Nopaline Synthase (NOS) polytermination sequence contained within the pBI-121 plasmid and contained a unique *Eco*RI restriction site to facilitate cloning of the vector into the pBI-121 binary vector backbone. The sense primers contained a unique *Bam*HI restriction site to facilitate cloning of the chimeric GUS reporter gene into the pBI-121 binary vector backbone. The primers used to generate each reporter are:

1. sense primer for pgus-IF (uidA-ATG-polyA-IF):  
5'- CCC GGA TCC ATG TTA AAA AAA AAA AAA AAA CGT CCT GTA GAA ACC-3' (SEQ ID NO:11)

2. sense primer for pgus-OF (uidA-ATG-polyA-OF):  
5'- CCC GGA TCC ATG TTA AAA AAA AAA AAA AAA ACG TCC TGT AGA AAC C-3' (SEQ

ID NO:12)

3. antisense primer (Nos-term) :

5'- CCC GAA TTC CCC GAT CTA GTA ACA TAG ATG-3' (SEQ ID NO:13)

PCR amplifications were carried out using reaction buffers described above.

Reactions were performed using 1 ng of pBI-121 vector as template (Life Technologies) and the appropriate corresponding primers. Amplifications were carried at 94°C for 30 seconds, 54°C for 60 seconds and 72°C for 60 seconds for 25 cycles. PCR products of the expected molecular weight was gel purified, cloned into T-tailed vectors (InVitrogen), and sequenced to ensure authentic sequence using the following primers: CaMV-FORW. [= 5'-gat atc tcc act gac gta ag-3'] (SEQ ID NO:30) for sequencing from the CaMV promoter into the 5' end of GUS cDNAs; NOSpA-42F [= 5'-tgt tgc cgg tct tgc gat g-3'] (SEQ ID NO:31) for sequencing of the NOS terminator; NOSpA-Cend-R [= 5'-ccc gat cta gta aca tag atg-3'] (SEQ ID NO:32) for sequencing from the NOS terminator into the 3' end of the GUS cDNAs; GUS-63F [= 5'-cag tct gga tcg cga aaa ctg-3'] (SEQ ID NO:33), GUS-441F [= 5'-ggt gat tac cga cga aaa cg-3'] (SEQ ID NO:34), GUS-825F [= 5'-agt gaa ggg cga aca gtt cc-3'] (SEQ ID NO:35), GUS-1224F [= 5'-gag tat tgc caa cga acc-3'] (SEQ ID NO:36), GUS-1596F [= 5'-gta tca ccg cgt ctt tga tc-3'] (SEQ ID NO:37), GUS-265R [= 5'-cga aac gca gca cga tac g-3'] (SEQ ID NO:38), GUS-646R [= 5'-gtt caa cgc tga cat cac c-3'] (SEQ ID NO:39), GUS-1033R [= 5'-cat gtt cat ctg ccc agt cg-3'] (SEQ ID NO:40), GUS-1425R [= 5'-gct ttg gac ata cca tcc-3'] (SEQ ID NO:41), and GUS-1783R [= 5'-cac cga agt tca tgc cag-3'] (SEQ ID NO:42) for the sequence of the full length GUS cDNAs. No mutation were found in either the OF or IF version of the GUS cDNA, and the expected frames for both cDNAs were also confirmed. pCR-IF-GUS and pCR-OF-GUS plasmids were subsequently digested with the BamH I and EcoR I restriction endonucleases, to generate DNA fragments containing the GUS cDNA along with the NOS terminator. These fragments were ligated into the BamH I and the EcoR I sites of the pBI-121 plasmid, which was prepared for cloning by cutting it with the same enzymes to release the wild type GUS cDNA. The resulting constructs (pBI-IF-GUS and pBI-OF-GUS) were subsequently digested with Hind III and EcoR I to release the DNA fragments encompassing the CaMV promoter, the IF or OF GUS cDNA, and the NOS terminator.

Finally, these fragments were ligated into the correspondent restriction sites present in the pGPTV-HPT binary vector (ATCC) to obtain the pCMV-IF-GUS-HPT and pCMV-OF-GUS-HPT binary vectors.

The resulting vectors, CMV-OF-GUS-HPT and CMV-IF-GUS-HPT now contain the CaMV35S promoter from the Cauliflower Mosaic 35 S Virus driving the GUS gene followed by a NOS terminator and polyadenylation signal (Fig. 7). In addition, this vector also contains a hygromycin resistance gene as a selectable marker that is used to select for plants containing this reporter.

#### **Generation of GUS reporter-expressing *Arabidopsis thaliana* transgenic plants.**

*Agrobacterium tumefaciens* bacteria are used to shuttle binary expression vectors into plants. To generate  $\beta$ -glucuronidase-expressing *Arabidopsis thaliana* (*A. thaliana*) plants, *Agrobacterium tumefaciens* cells (strain GV3101) were electroporated with the CMV-OF-GUS-HPT or the CMV-IF-GUS-HPT binary vector using methods known by those skilled in the art. Briefly, one-month old *A. thaliana* (ecotype Columbia) plants were infected by immersion in a solution containing 5% sucrose, 0.05% silwet and binary vector-transformed *Agrobacteria* cells for 10 seconds. These plants were then grown at 25°C under a 16 hour day and 8 hour dark photoperiod. After 4 weeks, seeds (referred to as T1) were harvested and dried for 5 days. Thirty thousands seeds from ten CMV-OF-GUS-HPT or CMV-IF-GUS-HPT-transformed plants were sown in solid Murashige and Skoog (MS) media plates in the presence of 20  $\mu$ g/ml of hygromycin (HYG). Three hundred plants were found to be HYG resistant and represented GUS expressing plants. These plants along with 300 control plants were grown in MS media for two weeks and then transferred to soil. Plants were grown for an additional four weeks under standard conditions at which time T2 seeds were harvested.

To confirm the integration and stability of the GUS vector in the plant genome, gene segregation and PCR analyses were conducted. Commonly, three out of four T1 plants transformed by *Agrobacteria* technology are expected to carry the vector inserted within a single locus and are therefore considered heterozygous for the integrated gene.

Approximately 75% of the seeds (T2) generated from most of the T1 plants were found HYG-resistant and this in accordance with the expected 1:2:1 ratio of null (no GUS

containing plants), heterozygous, and homozygous plants, respectively, in self-pollinating conditions. To confirm that these plants contained the GUS expression vector, genomic DNA was isolated from leaves of T1 plants using the DNAzol-mediated technique as described above. One ng of genomic DNA was analyzed by polymerase chain reaction (PCR) to confirm the presence of the GUS vector. PCR was carried out for 25 cycles with the following parameters: 95°C for 30 seconds; 54°C for 1 minute; and 72°C for 2 minutes using primers listed above. Positive reactions were observed in DNA from CMV-OF-GUS-HPT and CMV-IF-GUS-HPT-transformed plants and not from control (uninfected) plants.

In order to assess the expression of the GUS in T1 plants, leaf tissue was collected from T1 plants, homogenized in liquid nitrogen using glass pestles, and suspended in RLT lysing buffer (Qiagen, RNeasy plant RNA extraction kit). Five micrograms of total RNA was purified according to the manufacturer's suggested protocol and then loaded onto a 1.2% agarose gel (1x MOPS buffer, 3% formaldehyde), size-fractionated by electrophoresis, and transferred onto N-Hybrid+ membrane (Amersham). Each membrane was incubated at 55°C in 10 ml of hybridization solution (North2South labeling kit, Pierce) containing 100 ng of GUS, tubulin, or HYG probes, which were generated by PCR amplification, according to the manufacturer's directions. Membranes were washed three times in 2x SSC, 0.1% SDS at 55°C, and three times in 2x SSC at ambient temperature. Detection was carried out using enhanced chemiluminescence (ECL). GUS message was detected in three out of ten analyzed transgenic lines, while no signal was found in the control plants. Collectively these studies demonstrated the generation of GUS expressing transgenic *A. thaliana* plants.

To determine the status of MMR activity in host plants, one can measure for the production of functional  $\beta$ -glucuronidase by staining plant leaves or roots *in situ* for  $\beta$ -glu activity. Briefly, plant tissue is washed twice with water and fixed in 4 mls of 0.02% glutaraldehyde for 15 minutes. Next, tissue is rinsed with water and incubated in X-glu solution [0.1M NaPO<sub>4</sub>, 2.5 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 2.5mM K<sub>4</sub>Fe(CN)<sub>6</sub>, 1.5 mM MgCl<sub>2</sub>, and 1 mg/ml X-GLU (5 bromo-4-chloro-3-indoyl-  $\beta$ -D-glucuronide sodium salt) (Gold Biotechnology)] for 6 hours at 37°C. Tissues are then washed twice in phosphate buffered saline (PBS) solution, once in 70% ethanol and incubated for 4 hours in methanol:acetone (3:1) for 8 hours

to remove chlorophyll. Tissues are then washed twice in PBS and stored in PBS with 50% glycerol. Plant tissue with functional GUS activity will stain blue.

The presence of GUS activity in CMV-IF-GUS-HPT plants indicates that the in-frame N-terminus insertion of the poly A repeat does not disrupt the GUS protein function. The CMV-OF-GUS-HPT plants treated with DMA, EMS or untreated are tested to determine if these plants produce GUS activity. The presence of GUS activity in DMA treated plants indicates that the polyA repeat was altered, therefore, resulting in a frame-restoring mutation. Agents such as EMS, which are known to damage DNA by alkylation cannot affect the stability of a polynucleotide repeat. This data indicates that plants are defective for MMR, the only process known to be responsible for MI.

These data demonstrate the utility and power of using a chemical inhibitor of MMR to generate a high degree of genetic alteration that is not capable by means of standard DNA damaging drugs. Moreover, this application teaches of the use of reporter genes such as GUS-OF in plants to monitor for the MMR activity of a plant host.

#### **EXAMPLE 7: Use of chemical MMR inhibitors yields microsatellite instability in microbes.**

To demonstrate the ability of chemical inhibitors to block MMR in a wide range of hosts, we employed the use of *Pichia* yeast containing a pGUS-OF reporter system similar to that described in Example 5. Briefly, the GUS-OF and GUS-IF gene, which contains a polyA repeat at the N-terminus of the protein was subcloned from the pCR-IF-GUS and pCR-OF-GUS plasmids into the EcoRI site of the pGP vector, which is a constitutively expressed yeast vector containing a zeocin resistance gene as selectable marker. pGP-GUS-IF and pGP-GUS-OF vectors were electroporated into competent *Pichia* cells using standard methods known by those skilled in the art. Cells were plated on YPD agar (10g/L yeast extract; 20 g/L peptone; 2% glucose; 1.5% bactoagar) plates containing 100 µg/ml zeocin. Recombinant yeast are then analyzed for GUS expression/function by replica plating on YPD agar plates containing 100 µg/ml zeocin plus 1 mg/ml X-glu (5-bromo-4-chloro-3-indoyl-beta-D-glucuronide sodium salt) and grown at 30°C for 16 hours. On hundred percent of yeast expressing GUS-IF were found to turn blue in the presence of the X-glu substrate while none of the control yeast

turned blue. None of the yeast containing the GUS-OF turned blue in the presence of the X-glu substrate under normal growth conditions.

To demonstrate the ability of chemicals to block MMR in yeast, GUS-OF and control cells were incubated with 300 µM DMA, EMS, or no chemical for 48 hours. After incubation, yeast were plated on YPD-ZEO-X-GLU plates and grown at 30°C for 16 hours. After incubation, a subset of yeast expressing GUS-OF contain blue subclones, while none are seen in EMS or control cells. These data demonstrate the ability of chemicals to block MMR of microbes *in vivo* to produce subclones with new output traits.

EXAMPLE 8: Classes of other chemicals capable of blocking MMR *in vivo*

The discovery of anthracene compounds presents a new method for blocking MMR activity of host organisms *in vivo*. While 9,10-dimethylanthracene (DMA) was found to block MMR in cell hosts, other analogs with a similar chemical composition from this class are also claimed in this invention. These include anthracene and related analogs such as 9,10-diphenylanthracene and 9,10-di-M-tolylanthracene. Myers *et al.* ((1988) *Biochem. Biophys. Res. Commun.* 151:1441-1445) disclosed that at high concentrations, DMA acts as a potent weak mutagen, while metabolized forms of DMA are the "active" ingredients in promoting mutation. This finding suggests that metabolites of anthracene-based compounds may also act as active inhibitors of MMR *in vivo*. For instance, metabolism of anthracene and 9,10-dimethylanthracene by *Micrococcus sp.*, *Pseudomonas sp.* and *Bacillus macerans* microbes have found a number of anthracene and 9,10-dimethylanthracene metabolites are formed. These include anthracene and 9,10-dimethylanthracene cis-dihydrodiols, hydroxy-methyl-derivatives and various phenolic compounds. Bacteria metabolize hydrocarbons using the dioxygenase enzyme system, which differs from the mammalian cytochrome P-450 monooxygenase. These findings suggest the use of bacteria for biotransforming anthracene and DMA for additional MMR blocking compounds (Traczewska, T.M. *et al.* (1991) *Acta Microbiol. Pol.* 40:235-241). Metabolism studies of DMA by rat-liver microsomal preparations has found that this molecule is converted to 9-Hydroxymethyl-10-methylanthracene (9-OHMeMA) and 9,10-dihydroxymethyl-anthracene (9,10-DiOHMeA) (Lamparczyk, H.S. *et al.* (1984) *Carcinogenesis* 5:1405-1410). In addition, the trans-1,2-

dihydro-1,2-dihydroxy derivative of DMA (DMA 1,2-diol) was found to be a major metabolite as determined by chromatographic, ultraviolet (UV), nuclear magnetic resonance (NMR), and mass spectral properties. DMA 1,2-diol was also created through the oxidation of DMA in an ascorbic acid-ferrous sulfate-EDTA system. Other dihydriodols that are formed from DMA by metabolism are the trans-1,2- and 3,4-dihydriodols of 9-OHMeMA (9-OHMeMA 1,2-diol and 9-OHMeMA 3,4-diol) while the further metabolism of DMA 1,2-diol can yield both of these dihydriodols. Finally, when 9-OHMeMA is further metabolized, two main metabolites are formed; one was identified as 9,10-DiOHMeA and the other appeared to be 9-OHMeMA 3,4-diol.

The metabolism of 9-methylanthracene (9-MA), 9-hydroxymethylanthracene (9-OHMA), and 9,10-dimethylanthracene (9,10-DMA) by fungus also has been reported (Cerniglia, C.E. et al. (1990) *Appl. Environ. Microbiol.* 56:661-668). These compounds are also useful for generating DMA derivatives capable of blocking MMR. Compounds 9-MA and 9,10-DMA are metabolized by two pathways, one involving initial hydroxylation of the methyl group(s) and the other involving epoxidation of the 1,2- and 3,4- aromatic double bond positions, followed by enzymatic hydration to form hydroxymethyl trans-dihydriodols. For 9-MA metabolism, the major metabolites identified are trans-1,2-dihydro-1,2-dihydroxy and trans-3,4-dihydro-3,4-dihydroxy derivatives of 9-MA and 9-OHMA, whereby 9-OHMA can be further metabolized to trans-1,2- and 3,4-dihydriodol derivatives. Circular dichroism spectral analysis revealed that the major enantiomer for each dihydriodol was predominantly in the S,S configuration, in contrast to the predominantly R,R configuration of the trans-dihydriodol formed by mammalian enzyme systems. These results indicate that *Caenorhabditis elegans* metabolizes methylated anthracenes in a highly stereoselective manner that is different from that reported for rat liver microsomes.

The analogs as listed above provide an example but are not limited to anthracene-derived compounds capable of eliciting MMR blockade. Additional analogs that are of potential use for blocking MMR are shown in Fig.8.

**Other classes of small molecular weight compounds that are capable of blocking MMR *in vivo*.**

MMR is a multi-step process that involves the formation of protein complexes that detect mismatched bases or altered repetitive sequences and interface these mutations with enzymes that degrade the mutant base and repair the DNA with correct nucleotides. First, mismatched DNA is recognized by the mutS heterodimeric complex consisting of MSH2 and GTBP proteins. The DNA bound mutS complex is then recognized by the mutL heterodimeric complex that consists of PMS2 and MLH1 proteins. The mutL complex is thought to interface exonucleases with the mismatched DNA site, thus initiating this specialized DNA repair process. After the mismatched bases are removed, the DNA is repaired with a polymerase.

There are several steps in the normal process that can be targeted by small molecular weight compounds to block MMR. This application teaches of these steps and the types of compounds that may be used to block this process.

#### **ATPase inhibitors:**

The finding that nonhydrolyzable forms of ATP are able to suppress MMR *in vitro* also suggest that the use for this type of compound can lead to blockade of MMR *in vivo* and mutation a host organism's genome (Galio, L. *et al.* (1999) *Nucl. Acids Res.* 27:2325-2331; Allen, D.J. *et al.* (1997) *EMBO J.* 16:4467-4476; Bjornson, K.P. *et al.* (2000) *Biochem.* 39:3176-3183). One can use a variety of screening methods described within this application to identify ATP analogs that block the ATP-dependent steps of mismatch repair *in vivo*.

#### **Nuclease inhibitors:**

The removal of mismatched bases is a required step for effective MMR (Harfe, B.D. and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399). This suggests that compounds capable of blocking this step can lead to blockade of MMR *in vivo* and mutation a host organism's genome. One can use a variety of screening methods described within this application to identify nuclease inhibitors analogs that block the nuclease steps of mismatch repair *in vivo*. An example of the types of nuclease inhibitors are but not limited to analogs of N-Ethylmaleimide, an endonuclease inhibitor (Huang, Y.C., et.al. (1995) *Arch. Biochem. Biophys.* 316:485), heterodimeric adenine-chain-acridine compounds, exonulcease III

inhibitors (Belmont P, et.al., *Bioorg Med Chem Lett* (2000) 10:293-295), as well as antibiotic compounds such as Heliquinomycin, which have helicase inhibitory activity (Chino, M, et.al. *J. Antibiot. (Tokyo)* (1998) 51:480-486).

### Polymerase inhibitors:

Short and long patch repair is a required step for effective MMR (Modrich, P. (1994) *Science* 266:1959-1960). This suggests that compounds capable of blocking MMR-associated polymerization can lead to blockade of MMR *in vivo* and mutation a host organism's genome. One can use a variety of screening methods described within this application to identify polymerase inhibitors analogs that block the polymerization steps of mismatch repair *in vivo*. An example of DNA polymerase inhibitors that are useful in blocking MMR activity include, but are not limited to, analogs of actinomycin D (Martin, S.J., et.al. (1990) *J. Immunol.* 145:1859), Aphidicolin (Kuwakado, K. et.al. (1993) *Biochem. Pharmacol.* 46:1909) 1-(2'-Deoxy-2'-fluoro-beta-L-arabinofuranosyl)-5-methyluracil (L-FMAU) (Kukhanova M, et.al., *Biochem Pharmacol* (1998) 55:1181-1187), and 2',3'-dideoxyribonucleoside 5'-triphosphates (ddNTPs) (Ono, K., et.al., *Biomed Pharmacother* (1984) 38:382-389).

### Chemical Inhibitors of Mismatch Repair Gene Expression

MMR is a multi-protein process that requires the cooperation of several proteins such as but not limited to mutS homologs, MSH2, MSH3, MSH6, GTBP; mutL homologs PMS1, PMS2, MLH1; and exonucleases and helicases such as MutH and MutY (Harfe, B.D. and S. Jinks-Robertson (2000) *Ann. Rev. Genet.* 34:359-399; Modrich, P. (1994) *Science* 266:1959-1960). Chemicals capable of blocking the expression of these genes can lead to the blockade of MMR. An example of a chemical that is capable of blocking MMR gene expression is an oligodeoxynucleotide that can specifically bind and degrade an MMR gene message and protein production as described by Chauhan DP, et.al. (*Clin Cancer Res* (2000) 6:3827-3831). One can use a variety of screening methods described within this application to identify inhibitors that block the expression and/or function of MMR genes *in vivo*.

## DISCUSSION

The results described herein demonstrate the use of chemicals that can block mismatch repair of host organisms *in vivo* to produce genetic mutations. The results also demonstrate the use of reporter systems in host cells and organisms that are useful for screening chemicals capable of blocking MMR of the host organism. Moreover, the results demonstrate the use of chemical inhibitors to block MMR in mammalian cells, microbes, and plants to produce organisms with new output traits. The data presented herein provide novel approaches for producing genetically altered plants, microbes, and mammalian cells with output traits for commercial applications by inhibiting MMR with chemicals. This approach gives advantages over others that require the use of recombinant techniques to block MMR or to produce new output traits by expression of a foreign gene. This method will be useful in producing genetically altered host organisms for agricultural, chemical manufacturing, pharmaceutical, and environmental applications.

## PMS2 (mouse) (SEQ ID NO:14)

MEQTEGVSTE	CAKAIPIDG	KSVHQICSGQ	VISSLSTAVK	ELIENSVDAG	ATTIDLRLKD	60
YGVDLIEVSD	NGCGVEEENF	EGLALKHHTS	KIQEFADLTQ	VETFGFRGEA	LSSLCALSDV	120
TISTCHGSAS	VGTRLVFDHN	GKITQKTPYP	RPKGTTVSVQ	HLFYTLPVRY	KEFQRNIKKE	180
YSKMQVQLQA	YCIISAGVRV	SCTNQLGQK	RHAVVCTSQT	SGMKENIGSV	FGQKQLQSLI	240
PFVQLPPSDA	VCEYGLSTS	GRHKTFSTFR	ASFHSARTAP	GGVQQTGSFS	SSIRGPVTQQ	300
RSLSLSMRFY	HMYNRHQYPF	VVLNVSDSE	CVDINVTPDK	RQILLQEEKL	LLAVLKTSLI	360
GMFDSDANKL	NVNQQPLLDV	EGNLVKLHTA	ELEKPVPGKQ	DNSPSLKSTA	DEKRVASISR	420
LREAFSLHPT	KEIKSRGPET	AELTRSFPE	KRGVLSSYPS	DVISYRGLRG	SQDKLVSPTD	480
SPGDCMDREK	IEKDSGLSST	SAGSEEEFST	PEVASSFSSD	YNVSSLEDRP	SQETINCSDL	540
DCRPPGTGQS	LKPEDHGYQC	KALPLARLSP	TNAKRFKTEE	RPSNVNISQR	LPGPQSTSAA	600
EVDVAIKMNK	RIVILLEFSLS	SLAKRMQLQ	HLKAQNKHQL	SYRKFRAKIC	PGENQAAEDE	660
LRKEISKSMF	AEMEILGQFN	LGFIVTKLKE	DLFLVDQHAA	DEKYNFEMLO	QHTVLQAQRL	720
ITPQTLNLTA	VNEAVLIENL	EIFRKNGFDF	VIDEDAPVTE	RAKLISLPTS	KNWTFGPQDI	780
DELIFMLSDS	PGVMCRPSRV	RQMFASRACR	K SVMIGTALN	ASEMKKLITH	MGEMDHDPWNC	840
PHGRPTMRHV	ANLDVISQN					859

## PMS2 (mouse cDNA) (SEQ ID NO:15)

gaattccgg	taacctgtcg	tcaggtAACG	aagaatttcc	atgggtcctg	gtatcattgg	aggagacaga	60
gtctttcccc	gagagcggca	ccgcaactct	cccgcggta	ctgtgactgg	aggagtctg	ctggagacgc	120
catccatgga	gcaaaccgaa	ggcgtgagta	cagaatgtgc	taaggccatc	aagcctattg	180	
atgggaagtc	agtccatcaa	atttgtctg	ggcaggtat	actcaigtta	agcacccgctg	240	
tgaaggagtt	gataaaaaat	agtgtagatg	ctggtgctac	tactattgt	ctaaggctta	300	
aagactatgg	ggtgacac	attgaattt	cagacaatgg	atgtgggtt	gaagaagaaa	360	
actttgaagg	tctagctctg	aaacatcaca	catctaagat	tcaagat	gcccacccca	420	
cgcagggtga	aaccttcggc	tttcgggggg	aaagctctgag	ctctctgtgt	gcactaagt	480	
atgtcactat	atctacactgc	cacgggtctg	caagcgttgg	gactcgactg	gtgtttgacc	540	
ataatggaa	aatcacccag	aaaactccct	accccccacc	taaaggaaacc	acagtca	600	
tgcagcactt	attttataca	ctaccgtgc	gttacaaaga	gtttcagagg	aacattaaaa	660	

P  
9  
2  
0  
2  
3  
5  
5  
-  
0  
4  
5  
0  
E

```

aggagtattc caaaatggtg caggtcttac aggctactg tatcatctca gcaggggtcc 780
gtgttaagctg cactaatcag ctggacagg ggaagcggca cgctgtgtg tgcacaagcg 840
gcacgtctgg catgaaggaa aatatcggt ctgtgtttgg ccagaagcag ttgcaaagcc 900
tcattccctt tgtagctgt cccccctagt acgctgtgtg tgaagagtac ggcctgagca 960
cttcaggacg ccacaaaacc tttctactt ttcgggcttc atttcacagt gcacgcacgg 1020
cgccgggagg agtgcacacg acaggcagtt ttcttcata aatcagaggg cctgtgaccc 1080
agcaaaaggctc tctaagctt tcaatgaggt ttatcacat gtataaccgg catcagtacc 1140
catttgcgt ccttaacgtt tccgttact cagaatgtgt ggatattaat gtaactccag 1200
ataaaaaggca aattctacta caagaagaga agctattgtt ggcgtttt aagacctcct 1260
tgataggaat gtttagactg gatgcaaaca agcttaatgtt caaccagcag ccactgctag 1320
atgttgaagg taacttagta aagctgcata ctgcagaact agaaaaaggctt gtgccaggaa 1380
agcaagataa ctctccctca ctgaagagca cagcagacga gaaaaggta gcatccatct 1440
ccaggcttag agaggcctt tcttttcatc ctactaaaga gatcaagtctt aggggtccag 1500
agactgtca actgacacgg agttttccat gttgagaaaaag gggcgttta tcctcttata 1560
cttcagacgt catcttttac agaggcctt gttgctcgca ggacaaattt. gtgagtccca 1620
cgagcagccc tggtgactgt atggacagag agaaaaatataaaaactca gggctcagca 1680
gcacctcagc tggctctgag gaaggttca gcaccccccaga agtggccagt agcttttagca 1740
gtgactataa cgtgagctt ctagaaagaca gacccttctca gaaaccata aactgtggtg 1800
acctggactt ccgtccttca ggtacaggac agtctttagaa gccagaagac catggatata 1860
aatgcaaaagc tctaccttca gctcgctgtt caccacaaa tgccaaagcgc ttcagacacag 1920
aggaaagacc ctcaaatgtc aacatttctc aaagatttgc tggtcctcag agcacctcag 1980
cagctgaggt cgatgttagcc ataaaaatgtt ataagagaat cgtgctcctc gagttctc 2040
tgagttctt agctaaggcgtt atgaaggactt tacagcacctt aaaggcgcag aacaaacatg 2100
aactgaggta cagaaaaattt agggccaaaga ttggccctgg agaaaaaccggcagcagaag 2160
atgaacttgcgaaaagatgaaatcga tggttgcaga gatggagatc ttgggtca 2220
ttaacctggg atttatgtt accaaacttgc aagaggacctt cttccctggg gaccagcat 2280
ctcgccatgttga gaagtacaac ttggatgtc tgccggcggca cacggcgtc caggcgcaga 2340
ggctcatcac accccagact ctgaaacttgc tggctgtcaa tgaagtgta ctgatagaaa 2400
atctggaaat attcagaaag aatggcttgc accttgcatttcat tgatggatgtc gtcaggatca 2460
ctgaaaggcc taaaatttgcatttcccttacca ctagtacaaa ctggacccctt ggaccccaag 2520
atatacatgttactgtatctt atgttaatgtt acagccctgg ggtcatgtc cggcccttac 2580
gagtcagaca gatgtttgtt tccagagcctt gtcggaaagtc agtgtatgtt ggaacggcgc 2640
tcaatgcgag cgagatgtt aagctcatca cccacatggg tgagatggac cacccttggg 2700
actgccccca cggcaggccca accatgaggc acgttgcacca tctggatgtc atctctcaga 2760
actgacacac cccttgcgttca atagaggtaa ttacagattt ttcgggttgc aaagagaagg 2820
ttttaaatgttactgttacca gttgtacaaa aatttagcatgtt ctgcttaat gtactggatc 2880
catttaaaatgttactgttacca gttgtacaaa aatttagcatgtt ctgcttaat gtactggatc 2940
tgatccgttgc gtagctcatgttactgttacca gttgtacaaa aatttagcatgtt ctgcttaat gtactggatc 3000
agactcaatttcaaggacaaa aaaaaaaaaaaaatgttacca gttgtacaaa aatttagcatgtt ctgcttaat gtactggatc 3056

```

**PMS2 (human) (SEQ ID NO:16)**

MERAESSSTE	PAKAIKPIDR	KSVHQICSGQ	VVLSLSTAVK	ELVENSLDA	ATNIDLKLKD	60
YGVLDIEVSD	NGCIVVEEENF	EGLTLKHHTS	KIQEFADLTQ	VETFGFRGEA	LSSLCALSDV	120
TISTCHASAK	VGTRLMFDHN	GKIIQKTPYP	RPRGTTVSVO	QLFSTLPVRH	KEFQRNIKKE	180
YAKMVQVLHA	YCIISAGIRV	SCTNQLGQKG	RQPVVCTGGG	PSIKNIGSV	FGQKQLQSLI	240
PFVQLPPSDS	VCEEYGLSCS	DALHNLFYIS	GFISQCTHGV	GRSSTDQFF	FINRRPCDPA	300
KVCRVLVNEVY	HMYNRHQYPF	VVLNISVDSE	CVDINVTPDK	RQILLQEEKL	LLAVLKTSLI	360
GMFDSDVNKL	NVSQQPLLDV	EGNLIMKMAA	DLEKPMVKEQ	DQSPSLRTGE	EKKDVSISRL	420
REAFSLRHTT	ENKPHSPKTP	EPRRSPLGOK	RGMLSSSTSG	AISDKGVLRP	QKEAVSSSHG	480
PSDPTDRAEV	EKDSGHGSTS	VDSEGFSIPD	TGSHCSSEYYA	ASSPGDRGSQ	EHVDSQEKA	540
ETDDSFSDVD	CHSNQEDTGC	KFRVLPQPTN	LATPNTKRKF	KEEILSSSDI	CQKLVNTQDM	600
SASQVDVAVK	INKKVVPDF	SMSSLAKRIK	QLHHEAQQSE	GEQNYRKFRA	KICPGENQAA	660
EDELRKEISK	TMFAEMEIIG	QFNLGFIITK	LNEDIFIVDQ	HATDEKYNFE	MLQQHTVLOQ	720
QRLIAPOTLN	LTAVNEAVLI	ENLEIFRKNG	FDFVIDENAP	VTERAKLISL	PTSKNWTFGP	780
QDVDELIFML	SDSPGVMCRP	SRVKQMFASR	ACRKSVMIGT	ALNTSEMKKL	ITHMGEMDHP	840
WNCPHGRPTM	RHIANLGVIS	QN				862

**PMS2 (human cDNA) (SEQ ID NO:17)**

```

cgaggccggat cgggtgttgc atccatggag cgagctgaga gctcgagtac agaacctgtt 60
aaggccatca aaccttgc tccatgttgc gttgttgc gtcggatgtt 120

```

ctgagtcataa gcactgcggta aaaggaggta gtagaaaaaca gtctggatgc tggtgccact 180  
 aatattgtatc taaagcttaa ggactatgga gtggatctta ttgaagtttca agacaatgga 240  
 tgtgggttag aagaagaaaa cttcgaaggc ttaactctga aacatcacac atctaagatt 300  
 caagagtttgc cgcacctaacc tcaggttgc acctttggct ttcggggggaa agctctgagc 360  
 tcactttgtc cactgagcga tgtcaccatt tctacctgcc acgcacatcgcc gaaggttggaa 420  
 actcgactga tgtttgcataa caatggaaa attatccaga aaaccccccata cccccggcccc 480  
 agagggacca cagtcagcgt gcagcagttt tttccacac tacctgtgcg ccataaggaa 540  
 tttcaaaagga atattaagaa ggagtatgcc aaaatggtcc aggtcttaca tgcataactgt 600  
 atcatttcag caggcatccg tgtaagttgc accaatcagc ttggacaagg aaaacgacag 660  
 cctgtggat gcacagggtgg aagccccagc ataaaggaaa atatcggttc tgggtttggg 720  
 cagaagcagt tgcaaaggcct cattcccttt gtcaagctgc ccccttagtga ctccgtgtgt 780  
 gaagagtacg gtttgcgtt ttcggatgtct ctgcataatc ttttttacat ctcagggttc 840  
 atttcacaaat gcacgcgttgc agttggaaagg agttcaacag acagacagttt tttttttatc 900  
 aaccggccgc cttgtgaccgc agcaaaaggc tgcaagactcg tgaatggatgtt ctaccacatg 960  
 tataatcgac accagtatcc atttgttgcataa cttacattt ctgttgcattt agaatgcgtt 1020  
 gatatcaatg ttactccaga taaaaggcaaa attttgcataa aagaggaaaaa gctttgttg 1080  
 gcagttttaa agacctctt gataggaatg tttgatagtg atgtcaacaa gctaaatgtc 1140  
 agtcagcagc cactgctggaa tgtaaggtt aacttaataaa aatgcacatgc agcggatttg 1200  
 gaaaagccca tggttagaaaaa gcaggatcaa tccccttcat taaggactgg agaagaaaaa 1260  
 aaagacgtgt ccatttccag actgcgagag gcttttctc ttcgtcacac aacagagaac 1320  
 aaggcctcaca gcccaaagac tccagaacca agaaggagcc ctctagggaca gaaaagggg 1380  
 atgctgtctt ctagcacttc aggtgcacatc tctgacaaag gcgtcctgag acctcagaaaa 1440  
 gaggcagtgaa gtttgcgtca cggaccaggat gaccctacgg acagagcgga ggtggagaag 1500  
 gactcggggc acggcagcac ttccgtggat tctgaggggt tcagcatccc agacacgggc 1560  
 agtcactgca gcagcggatc tgccgcgc tcccccagggg acaggggctc gcaggaacat 1620  
 gtggactctc aggagaaaaagc gcctgaaact gacgactctt ttccatgtt ggactgcct 1680  
 tcaaaccagg aagataccgg atgtaaattt cgagtttgc ctcaagccaaatc taatctcgca 1740  
 acccccaaca caaagcgttt taaaaaaagaa gaaattctt ccagttctga catttgcataa 1800  
 aagtttagtaa atactcagga catgtcagcc ttcagggtt atgttagctgt gaaaattaaat 1860  
 aagaaaagtt tgcccttggaa cttttctatg agttctttag cttaaacgaat aaagcagttt 1920  
 catcatgaag cacagcaaag tgaagggggaa cagaattaca ggaagtttag ggcaaaagatt 1980  
 tgcctctggag aaaatcaagc agccgaagat gaactaagaa aagagataag taaaacgatg 2040  
 tttgcagaaa tggaaatcat tggtcagttt aacctgggat ttataataac caaactgaat 2100  
 gaggatatct tcatagtggaa ccagcatgcc accgacgaga agtataactt cgagatgtc 2160  
 cagcagcaca ccgtctccca ggggcagagg ctcatagcac ctcagactct caacttaact 2220  
 gctgttaatg aagctgttct gatagaaaaat ctggaaatattt ttagaaagaa tggctttgtat 2280  
 tttgttatcg atgaaaatgc tccagtcact gaaaggcata aactgttttctt ctggccaact 2340  
 agtaaaaact ggaccttcggg accccaggac gtcgatgaac tgatcttcat gtcgagcgc 2400  
 agccctgggg tcatgtggcc gccttcccgaa gtciaagcaga tggttgcctc cagagcctgc 2460  
 cggaaagtcgg tgatgattgg gactgctt aacacaagcg agatgaagaa actgtatcacc 2520  
 cacatggggg agatggacca cccctggaaac tgcctccatg gaaggccaaac catgagacac 2580  
 atcgccaacc tgggtgtcat ttctcagaac tgaccgtatg cactgtatgg aataatttgtt 2640  
 tttatcgcag atttttatgt tttgaaagac agagtcttca ctaacctttt ttgttttaaa 2700  
 atgaaacactg ctacttaaaa aaaatacaca tcacacccat taaaagtga tcttgagaac 2760  
 ctttcaaac c 2771

## PMS1 (human) (SEQ ID NO:18)

MKQLPAATVR LLSSSQIITS VVSVVKELIE NSLDAGATSV DVKLENYGF D KIEVRDNNGEG 60  
 IKAVDAPVMA MKYYTSKINS HEDLENLTTY GFRGEALGSII CCIAEVLITT RTAADNFSTQ 120  
 YVLDGSGHIL SQKPSHLGQG TTVTALRLFK NLPVRKQFYS TAKKCKDEIK KIQDLMMSFG 180  
 ILKPDLRIVF VHNVKAVIWQK SRVSDHKMAL MSVLGTAVMN NMESFQYHSE ESQIYLSGFL 240  
 PKCDADHSFT SLSTPERSFI FINSRPVHQK DILKLIRHHY NLKCLKESTR LYPVFFLKID 300  
 VPTADVDVNL TPDKSQVLLQ NKESVLIALE NLMTTCYGPL PSTNSYENNKT DVSAADIVL 360  
 SKTAETDVLF NKVESSGKNY SNVDTSVIPF QNDMHNDESG KNTDDCLNHQ ISIGDFGYGH 420  
 CSSEISNIIDK NTKNAFQDIS MSNVSWENSQ TEYSKTCFIS SVKHTQSENG NKDHIDESGE 480  
 NEEEAGLENS SEISADEWSR GNILKNSVGE NIEPVKILVP EKSLPCKVSN NNYPipeQMN 540  
 LNEDSCNKKS NVIDNKGKV TAYDLLSNRV IKKPMASAL FVQDHGPQFL IENPKTSLED 600  
 ATLQIEELWK TLSEEELKLY EEKATKDLER YNSQMKRAIE QESQMSLKDG RKKIKPTEAW 660  
 NLAQKHKLKT SLSNQPKLDE LLQSQIEKRR SQNIKMQVQIP FSMKNLKINF KKQNQKVDLEE 720  
 KDEPCLIHNL RFPDAWLMTS KTEVMLNPY RVEEALLFKR LLENHKLPAE PLEKPIMLTE 780  
 SLFNGSHYLD VLYKMTADDQ RYSGSTYLSD PRLTANGFKI KLIPGV SITE NYLEIEGMAN 840

CLPFYGVADL KEILNAILNR NAKEVYECRP RKVISYLEGE AVRLSRQLPM YLSKEDIQDI 900  
IYRMKHQFGN EIKECVHGRP FFHHLTYLPE TT 932

## PMS1 (human) (SEQ ID NO:19)

ggcacgagt gctgcttgcg gctagttggat ggtaattgcc tgcctcgcc tagcagcaag 60  
ctgctctgtt aaaagcgaaa atgaaacaat tgcctcgcc aacagtcga ctccttcaa 120  
gttctcaaat catcaactcg gtggtcagtg ttgtaaaaga gcttattgaa aactccttgg 180  
atgctgtgc cacaagcgta gatgttaaac tggagaacta tggatttgat aaaattgagg 240  
tgcgagataa cggggagggt atcaaggctg ttgatgcacc tgtaatggca atgaagtact 300  
acacacctaaa aataaaatagt catgaagatc ttgaaaattt gacaacttac ggtttcgtg 360  
gagaagcctt ggggtcaatt tggatgtat ctgagggttt aattacaaca agaacggctg 420  
ctgataaattt tagcaccaggat tggatgtttag atggcgttgg ccacatactt tctcagaaac 480  
cttcacatct tggcaaggt acaactgtaa ctgctttaag attattaag aatctacctg 540  
taagaaagca gtttactca actgcaaaaa aatgtaaaga tgaaataaaa aagatccaag 600  
atctcctcat gagcttttgtt atccttaaac ctgacttaag gattgtctt gtacataaca 660  
aggcagttat ttggcagaaaa agcagagttt cagatcacaat gatggctctc atgtcagttc 720  
tggggactgc tggatgtttag aatatggaat cctttagtgc ccactctgaa gaatctcaga 780  
tttatctcag tggatttctt ccaaagtgtg atgcagacca ctcttcaact agtcttcaa 840  
caccagaaag aagtttcatc ttcataaaca gtcgaccagt acatcaaaaa gatatcttaa 900  
agttaatccg acatcattac aatctgaaat gcctaaaggaa atctactcgt ttgtatcctg 960  
ttttcttctt gaaaatcgat gttcctacag ctgatgttga tgtaaatttta acaccagata 1020  
aaagccaagt attattacaa aataaggaaat ctgttttaat tgctcttgc aatctgtat 1080  
cgacttgc tggaccattt ccttagtacaa atctttatgaa aaataataaa acagatgttt 1140  
ccgcagctga catcgttctt agtaaaacacg cagaaacaga tggctttt aataaaagtgg 1200  
aatcatctgg aaagaatttat tcaaattgtt atacttcgtt cattccattt caaaatgata 1260  
tgcataatga tgaatctgga aaaaacactg atgattgtttt aaatcaccag ataagtattt 1320  
gtgactttgg ttatggcat tggatgtt aatattctaa cattgataaa aacactaaga 1380  
atgcatttca ggacatttca atgagtaatg tatcatggga gaactctcag acggaatata 1440  
gtaaaacttg tttataagt tccgttaagc acacccagtc agaaaatggc aataaaagacc 1500  
atatacatgtt gaggggaa aatgaggaag aagcaggtt tggaaactct tcggaaattt 1560  
ctgcagatgtt gtggagcagg gggaaatatac ttaaaaattt aatgggagag aatattgtt 1620  
ctgtgaaaat tttatgtcctt caatccctga acaaataatgaaat cttatgtt aatatttatac 1680  
ataataaaatc tggaaaatgtt acatgtttagt caatcgatg atcaagaaac 1740  
ccatgtcagc aagtgttctt tttgttcaag atcatcgatc ttagttctc atagaaaatc 1800  
ctaagacttag tttatgtttagt gcaacactac aatattgtt aatgttgcg aatattgtt 1860  
aagagaaaaa actgaaatatac aagagaaagg ctactaaaga cttggaaacga tacaatgtc 1920  
aaatgttgcg agccatttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg 1980  
taaaaaccac cagcgcatgg aatattggccc aagagcacaat gttaaaaacc tcattatcta 2100  
atcaacccaaa acttgcatttgc ttccttcattt cccaaatttgc aaaaagaagg agtcaaaaata 2160  
ttaaaatgtt acagatcccc acaaagtgttgc tttatgtt aaaaacttataa aataattttt aagaaacaaa 2220  
acaatccctga cttagaaagag aaggatgttgc tttatgttgc ccacaatctc aggtttcctg 2280  
atgcattgtt aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2340  
aagccctgtt attttttttt aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2400  
agccaaatatttgc tttatgttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2460  
aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2520  
cagcgatgg tttatgttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2580  
aaatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2640  
ttatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2700  
taagttttt aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2760  
aagaggacatttgc tttatgttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2820  
agtgtgttgc tttatgttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2880  
taatatgtt taagaagatttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg 2940  
tctggtttgc aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 3000  
ctgacttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg aatgttgcg 3060  
aac

## MSH2 (human) (SEQ ID NO:20)

MAVQPETLQ LESAAEVGFV RFFQGMPEKP TTVRLFDRG DFYTAHGEDA LLAAREVFKT 60

QGVIKYMGPA	GAKNLQSVVL	SKMNFEFVK	DLLLVRQYRV	EVYKNRAGNK	ASKENDWYLA	120
YKASPGNLSQ	FEDILFGNNND	MSASIGVVGV	KMSAVDGQRQ	VGVGVDSIQ	RKLGLCEFDP	180
NDQFSNLEAL	LIQIGPKECV	LPGGETAGDM	GKLRQIIQRG	GILITERKKA	DFSTKDIYQD	240
LNRLLGKKG	EQMNSAVLPE	MENQVAVSSL	SAVIKFLELL	SDDSNFGQFE	LTTFDFSQYM	300
KLDIAAVRAL	NLFQGSVEDT	TGSQSLAALL	NKCKTPQGQR	LVNQWIQPL	MDKNRIEERL	360
NLVEAFVEDA	ELRQTLQEDL	LRRFPDLNRL	AKKFQRQAAN	LQDCYRLYQG	INQLPNVIQA	420
LEKHEGKHQK	LLLAVFVTPL	TDLRSDFSKF	QEMIETTLD	DQVENHEFLV	KPSFDPNLSE	480
LREIMNDLEK	KMQSTLISAA	RDLGLDPGKQ	IKLDSSAQFG	YYFRVTCKEE	KVLRNNKNFS	540
TVDIQKNGVK	FTNSKLTSLN	EEYTKNTEY	EEAQDAIVKE	IVNISSGYVE	PMQTLNDVLA	600
QLDAVVSFAH	VSNGAPVPYV	RPAILEKGQG	RIILKASRHA	CVEVQDEIAF	IPNDVYFEKD	660
KQMFHIITGP	NMGGKSTYIR	QTGVIVLMAQ	IGCFVPCESA	EVSVDCILA	RVGAGDSQLK	720
GVSTFMAEML	ETASILRSAT	KDSLIIIDEL	GRGTSTYDF	GLAWAISEYI	ATKIGAFCMF	780
ATHFHELTAL	ANQIPTVNNL	HVTALTTEET	LTMLYQVKKG	VCDQSGFIHV	AELANFPKHV	840
IECAKQKALE	LEEFQYIGES	QGYDIMEPAA	KKCYLEREQG	EKIIQEFLSK	VKQMPFTEMS	900
EENITIKLKQ	LKAEVIAKNN	SFVNEIISRI	KVTT			934

## MSH2 (human cDNA) (SEQ ID NO:21)

ggccggaaac agcttagtgg gtgtgggtc ggcattttc ttcaaccagg aggtgaggag 60  
 gtttcgacat ggcgtgcag ccgaaggaga cgctgcagtt ggagagcgcg gcccaggatcg 120  
 gcttcgtcg cttcttcag ggcattccgg agaagccgac caccacagtg cgccttttcg 180  
 accggggcga ctctatacg gcgcacggcg agacgcgcg gctggccgcc cgggagggtgt 240  
 tcaagaccga gggggtgatc aagtacatgg ggcggcagg agcaaagaat ctgcagatg 300  
 ttgtgcttag taaaatgaat tttgaatctt ttgtaaaaga tcttcttctg gttcgtca 360  
 atagagttg agttataag aatagagctg gaataaaggc atccaaggag aatgattgt 420  
 atttgcata taaggcttc cctggcaatc tctctca 480  
 acaatgatat gtcagcttc atttgttgc tggtgttta aatgtccgc gttgatggcc 540  
 agagacagg tggagttggg tatgtggatt ccatacagag gaaactagga ctgtgtgaat 600  
 tccctgataa tgatcagttc tccaaatctt aggtcttc 660  
 aatgtgttt accccggagga gagactgctg gagacatggg gaaactgaga cagataattc 720  
 aaagaggagg aattctgatc acagaaagaa aaaaagctga cttttccaca aaagacattt 780  
 atcaggacct caaccgggtg ttgaaaggca aaaaggaga gcagatgaat agtgcgttat 840  
 tgccagaat ggagaatcag gttcagttt catcaactgtc tgcggtaatc aagtttttag 900  
 aacttcttac agatgattcc aactttggac agttgaact gactacttt gacttcagcc 960  
 agtatatgaa attgatatt gcapcactca gagcccttaa ctttttcag gttctgttg 1020  
 aagataccac tggctctcg tctctggctg cttgtctgaa taagtgtaaa acccctcaag 1080  
 gacaaagact ttttaaccag tggattaagc agcctctcat ggataagaac agaatagagg 1140  
 agagattgaa ttttagtgaa gctttttag aagatgcaga attgaggcag actttacaag 1200  
 aagatttact tcgtcgattt ccagatctt accgactgtc caagaatgtt caaagacaag 1260  
 cagcaaactt acaagattgt tacggactct atcagggtat aaatcaacta ctaatgtta 1320  
 tacaggctt gaaaaaacat gaaggaaaac accagaaatt attgtggca gtttttgtga 1380  
 ctcctcttac tgatcttcgt tctgacttct ccaagttca gaaaatgata gaaacaactt 1440  
 tagatatgga tcaggtggaa aaccatgaat tccttgtaaa accttcattt gatcctaattc 1500  
 tcagtgaatt aagagaaata atgaatgact tgaaaagaa gatgcgtca acattaataa 1560  
 gtgcagccag agatcttggc ttggaccctg gcaaacagat taaaactggat tccagtgcac 1620  
 agtttggata ttactttcgta gtaacctgta aggaagaaaa agtccttcgt aacaataaaa 1680  
 acttttagtac ttttagatatac cagaagaat gtgttaattt taccaacacg aaattgactt 1740  
 cttaaatgaa agatatacc aaaaataaaa cagaatatga agaagccccag gatgccattt 1800  
 taaaagaaat tttcatttctt tttcatttctt atttagaacc aatgcagaca ctaatgtat 1860  
 ttttagctca gctagatgtc ttgttcgt gtcattatgaa gcaatgtt aaaaatccca 1920  
 catatgtacg accagccatt ttggagaaag gacaaggaag aattatatta aaaaatccca 1980  
 ggcattctt gtttggatc caagatggaa ttgcattttt tcctaatgac gtatactttt 2040  
 aaaaagataa acagatgtt cacatcatta ctggcccaa tatggaggt aatcaacat 2100  
 atattcgaca aactgggtg atagtactca tggccaaat tgggtttt gttccatgt 2160  
 agtcagcaga agtgcattt gttggactgca tcttagcccg agtagggct ggtgacagtc 2220  
 aattgaaagg agtctccacg ttcatggctg aatgttgaa aactgcttctt atcctcagg 2280  
 ctgcacccaa agattcatta ataatcatag atgaattggg aagaggaact tctacctacg 2340  
 atggattttgg gtttagcatgg gctatatacg aatacattgc aacaaagatt ggtgctttt 2400  
 gcatgtttgc aaccattttt catgaactta ctgccttggc caatcagata ccaactgtta 2460  
 ataatctaca tgtcacagca ctcaccactt aagagacattt aactatgtt tatcaggtt 2520

agaaagggtgt ctgtgatcaa agttttggga ttcatgttgc agagcttgct aatttcccta 2580  
 agcatgtaat agagtgtgtc aaacagaaaag ccctggaact tgaggagttt cagtatattg 2640  
 gagaatcgca aggatatgt atcatgaaac cagcagcaaa gaagtgcstat ctggaaagag 2700  
 agcaagggtga aaaaattatt caggaggccc tggccaaggt gaaacaatg cccttactg 2760  
 aaatgtcaga agaaaacatc acaataaaagt taaaacagct aaaagctgaa gtaatagcaa 2820  
 agaataatag cttgttaaat gaaatcattt cacgaaataaa agtactacg tgaaaaatcc 2880  
 cagtaatggaa atgaaggtaa tattgataag ctattgtctg taatagttt atattgttt 2940  
 atattaaccc ttttccata gtgttaactg tcagtgccta tgggctatca acttaataag 3000  
 atatttagta atatttact ttgaggacat ttcaaaagat ttttattttg aaaaatgaga 3060  
 gctgttaactg aggactgtt gcaattgaca taggcaataa taagtgatgt gctgaatttt 3120  
 ataaataaaaa tcagtgtagtt tgtgg 3145

## MLH1 (human) (SEQ ID NO:22)

MSFVAGVIRR	LDETVVNRIA	AGEVIQRPAN	AIKEMIENCL	DAKSTSIQVI	VKEGGLKLIQ	60
IQDNGTGIK	EDDIVCERF	TTSKLQSFED	LASISTYGF	GEALASISHV	AHVTITTKTA	120
DGKACRYASY	SDGKLKAPPK	PCAGNQGTOI	TVEDLFYNIA	TRRKALKNPS	EYKGKILEVV	180
GRYSVHNAGI	SFSVKKQGET	VADVRTLPNA	STVDNRISIF	GNAVSRRELIE	IGCEDKTLAF	240
KMNGYISMAN	YSVKKCIFLL	FINHRLVEST	SLRKAIETVY	AAYLPKNTHP	FLYLSLEISP	300
QNVDVNVHPT	KHEVHFLHEE	SILERVQQHI	ESKLIGSNSS	RMYFTQTLLP	GLAGPSGEMV	360
KSTTSLTSSS	TSGSSDKVYA	HQMVRTDSRE	QKLDNFLQPL	SKPLSSQPQA	IVTEDKTDIS	420
SGRARQDDEE	MLELPAPAEV	AAKNQSLEGD	TTKGTSEMSE	KRGPTSSNPR	KRHREDSDVE	480
MVEDDSRKEM	TAACTPERRRI	INLTSVLSQ	EEINEQGHEV	LREMLHNHSF	VGCVNPQWAL	540
AQHQTKLYLL	NTTKLSEELF	YQILYYDFAN	FGVRLLSEPA	PLFDLAMLAL	DSPESGWTEE	600
DGPKEGLAHEY	IVEFLKKKAE	MLADYFSLEI	DEEGNLIGLP	LLIDNYVPPL	EGLPIFILRL	660
ATEVNWDEEK	ECFESLSKEC	AMFYSIRKQY	ISEESTLSQ	QSEVPGSIPN	SWKWTVEHIV	720
YKALRSHILP	PKHFTEDGNI	LQLANLPDLY	KVFERC			756

## MLH1 (human) (SEQ ID NO:23)

cttggcttt ctggcgc当地 aatgtcggtc gtggcagggg ttattcggcg gctggacgag 60  
 acagtggta accgcacatcg ggcggggaa gtatccagc ggccagctaa tgctatcaa 120  
 gagatgattt agaactgtt agatgcaaaa tccacaagta ttcaagtgtat ttttaaagag 180  
 ggaggccctga agttgattca gatccaagac aatggcaccc ggttcaggaa agaagatctg 240  
 gatattgtat gtggaaaggtt cactactgt aaactgcagt cctttgggaa tttagccagt 300  
 atttctacct atggcttcg aggtgaggct ttggccagca taagccatgt ggctcatgtt 360  
 actattacaa cgaaaaacagc tgatggaaag tgatccataca gagcaatgtt ctcagatgg 420  
 aaactgaaag cccctctaa accatgtgtc ggcaatcaag ggaccaggat cacgggtgg 480  
 gacctttt acaacatagc cacgaggaga aaagctttaa aaaatccaag tgaagaat 540  
 gggaaaattt tggaagttgt tggcaggat tatgtacaca atgcaggcat tagttctca 600  
 gttaaaaaac aaggagagac agtagctgtat gtaggtacac taccaatgc ctcaccgtg 660  
 gacaatattc gctccatctt tgaaatgtct gtagtcgag aactgataga aattggatgt 720  
 gaggataaaa ccctagcctt caaaatgtat gttacatata ccaatgcaaa ctactcgtg 780  
 aagaagtgc tcttcttact cttcatcaac catcgctcgg tagaatcaac ttccttgaga 840  
 aaagccatag aaacagtgtc tgcagcttat ttgccccaaa acacacaccc attcctgtac 900  
 ctcagtttag aaatcagtcc ccagaatgtc gatgttataat tgcacccac aaagcatgaa 960  
 gttcaacttcc tgcacgagga gaggatcgtc gagcgggtgc agcagcacat cgagagcaag 1020  
 ctcctgggt cccatcttc caggatgtac ttccatcaggat ctttgcatttcc aggttgtct 1080  
 gccccctctg gggagatggtaa taaatccaca acaagtgtc ctcgtcttc tacttctgaa 1140  
 agtagtgata aggtctatgc ccaccagatg gttcgtacat attccctggaa acagaagctt 1200  
 gatgcatttc tgcacgctct gacaaaccc ctgtccagtc agccccaggc cattgtcaca 1260  
 gaggataaga cagatatttca tagtggcagg gctaggcagc aagatgagga gatgtttgaa 1320  
 ctcccagccc ctgtgttgtt ggctgcaaaa aatcagatgtc tgggggggaa tacaacaaag 1380  
 gggacttcag aaatgtcaga gaagagagga cctacttcca gcaacccacaa agagacat 1440  
 cggaaagatt ctgtgtggaa aatgggtggaa gatgttccc gaaaggaaat gactgcagct 1500  
 tgcacccccc ggagaaggat cattaaccc actagtgttt tgagtcctca ggaagaaatt 1560  
 aatgagcagg gacatggatgt tctccgggag atgttgcata accactccctt cgtggctgt 1620  
 gtgaatccctc agtggccctt ggcacacat caaaccaggat tatacccttcc caacaccacc 1680  
 aagcttagt gaaactgtt ctacccatata ctcattatgt atttggccaa ttttgggttt 1740  
 ctcaggttat cggagccagc accgctctt gaccccttgc gaccccttgcata agatgttca 1800

MOR-0017

PATENT

gagagtggct ggacagagga agatggtccc aaagaaggac ttgctgaata cattgttgag 1860
tttctgaaga agaaggctga gatgctgca gactattct ctttggaaat tgatgaggaa 1920
gggaacctga ttgattacc cttctgatt gacaactatg tgccccctt ggagggactg 1980
cctatctca ttcttcgact agccacttagt gtgaattggg acgaagaaaa ggaatgtttt 2040
gaaaacctca gtaaagaatg cgctatgtt tattccatcc gaaagcaga catatctgag 2100
gagtcgaccc tctcaggcca gcagagtcaa gtgcctggct ccattccaaa ctcctggaaag 2160
tggactgtgg aacacattgt ctataaagcc ttgcgctcac acattctgcc tcctaaacat 2220
ttcacagaag atgaaatat cctgcagctt gctaaccctgc ctgatctata caaagtcttt 2280
gagaggtgtt aaatatggtt atttatgcac tggatgtt gttctcttt ctctgtattc 2340
cgatacaaaag tggatgtatca aagtgtgata tacaaaagtgt accaacataa gtgttggtag 2400
cacttaagac ttatacttgc cttctgatag tattcctta tacacagtgg attgattata 2460
aataaaataga tgtgtcttaa cata 2484

**hPMS2-134 (human) (SEQ ID NO:24)**

MERAESSSTE PAKAIKPIDR KSVHQICSGQ VVLSLSTAVK ELVENSLDAK ATNIDLKLKD 60
YGVDLIEVSD NGCGVEEENF EGLTLKHHTS KIQEFADLTQ VETFGFRGEA LSSLCALSDV 120
TISTCHASAK VGT 133

**hPMS2-134 (human cDNA) (SEQ ID NO:25)**

cgaggcggat cgggtgttgc atccatggag cgagctgaga gctcgagttac agaacctgtct 60
aaggccatca aacctattga tcggaagtca gtcacatcaga tttgctctgg gcagggggta 120
ctgagtcataa gcactgcgggt aaaggagttt gtagaaaaca gtctggatgc tgggtccact 180
aatattgtatc taaagcttaa ggactatggta gtggatctta ttgaagtttc agacaaatgg 240
tgtggggtag aagaagaaaa cttcgaaggc ttaactctgt aacatcacac atctaagatt 300
caagagttt cccacctaacc tcaggttggaa actttttggct ttcggggggaa agctctgagc 360
tcactttgtt cactgagcga tgcaccatt tctacactgccc acgcacatcggc gaagggttgg 420
acttga 426

**GTBP (human) (SEQ ID NO:26)**

MSRQSTLYSF FPKSPALSDA NKASARASRE GGAAAAPGA SPSPGGDAAW SEAGPGPRPL 60
ARSASPPKAK NLNGGLRRSV APAAPTSCDF SPGDLVWAKM EGYPWWPCLV YNHFDGTFI 120
REKGKSVRVH VQFDDSPTR GWVSKRLLKP YTGSKSKEAQ KGGHFYSAKP EILRAMQRAD 180
EALNKDKIKR LELAVCDEPS EPEEEEEEVE GTTYVTDKSE EDNEIESEEE VQPKTQGSRR 240
SSRQIKKRRV ISDSES DIGG SDVEFKPDTK EEGSSDEISS GVGDSSE EGL NSPVVKVARKR 300
KRMVTGNGL KRKSSRKETP SATKQATSIS SETKNTLRAF SAPQNSESQA HVSGGGDDSS 360
RPTVWYHETL EWLKEEKRRD EHRRRPDHDP DFASTLYVPE DFLNSCTPGM RKWWQIKSQN 420
FDLVICYKVG KFYELYHMDA LIGVSELGLV FMKGNWAHSG FPEIAFGRYS DSLVQKGYKV 480
ARVEQTETPE MMEARCRKMA HISKYDRVVR REICRIITKG TQTYSVLEGD PSENYSKYLL 540
SLKEKEEDSS GHTRAYGVCF VDTSLGKFFI GQFSDDRHC SRFRTLVAHYP PVQVLFEKGN 600
LSKETKTILK SSLSCSLQEG LIPGSQFWDA SKTLRTLLEE EYFREKLSDG IGVMLPQVLK 660
GMTSESDSIG LTPGEKSEL A LSALGGCFY LKTKLIDQEL LSMANFEYYI PLDSDTVSTT 720
RSGAIFTKAY QRMVLDATL NNLEIFLNGT NGSTEGTLL RVDTCHTPFG KRLLKQWLCA 780
PLCNHYAIND RLDAIEDLMV VPDKISEVVE LLKKLPDLER LLSKIHNVGS PLKSQNHPD 840
RAIMYEETTY SKKKIIDFLS ALEGFKVMCK IIIGIMEEVAD GFKSKILKQV ISLQTKNPEG 900
RFPDLTVELN RWDTAFDHEK ARKTGLITPK AGFDSDYDQA LADIRENEQS LLEYLEKQRN 960
RIGCRTIVYW GIGRNRYQLE IPENFTTRNL PEYEYLKSTK KGCKRYWTKT IEKKLANLIN 1020
AEERRDVSLK DCMMRLFYNF DKNYKDQSA VECIAVLDVL LCLANYSRGG DGPMCRPVIL 1080
LPEDTPPFLE LKGSRHPCIT KTFFGDDFIP NDILIGCEEE EQENGKAYCV LVTGPNMGGK 1140
STLMRQAGLL AVMAQMGCYV PAEVCRLTPI DRVFTRLGAS DRIMSGESTF FVELSETASI 1200
LMHATAHSLV LVDELGRGTA TFDGTAIANA VVKELAETIK CRTLFS THYH SLVEDYSQNV 1260
AVRLGHMACM VENECEDPSQ ETITFLYKFI KGACP KSYGF NAARLANLPE EVIOKGHRKA 1320
REFEKMNQSL RLFREVCLAS ERSTVDAEAV HKLLTLLIKE 1360

**GTBP (human cDNA) (SEQ ID NO:27)**

gccgcgcggc agatgcgggtt cttttaggag ctccgtccga cagaacgggtt gggccttgc 60
-----------------------------------------------------------------------

ggctgtcgg	atgtcgac	agagcaccc	gtacagctt	ttccccaagt	ctccggcgct	120
gagtgtgcc	aacaaggcct	cggccagggc	ctcacgcgaa	ggcgccgtg	ccgcgcgtc	180
ccccggggc	tctccttcc	caggcgggga	tgcggcttg	agcgaggctg	ggcctggcc	240
caggccctg	ggcgcgtcg	cgtcaccgc	caaggcgaag	aacctcaacg	gagggctgctg	300
gagatcgta	gcccgtctg	cccccaccag	ttgtgacttc	tcaccaggag	atttggttt	360
ggccaagatg	gagggttacc	cctggggcc	ttgtctggg	tacaaccacc	cctttagatgg	420
aacattcatc	cgcgagaaag	ggaaatcagt	ccgtgttcat	gtacagttt	ttgatgacag	480
cccaacaagg	ggctgggta	gcaaaaggct	ttaaagcca	tatacagggt	caaatactaa	540
ggaagccag	aagggggtc	attttacag	tgc当地	gaaatactga	gagaatgca	600
acgtgcagat	gaagcctta	ataaagacaa	gattaaggag	cttgaatttg	cagttgtga	660
tgagccctca	gagccagaag	aggaagaaga	gatggaggt	ggcacaactt	acgtaacaga	720
taagagtga	gaagataatg	aaatttggag	tgaagaggg	gtacagctt	agacacaagg	780
atctagggcga	atgtcgcc	aaataaaaaa	acgaagggtc	atatacgatt	ctgagagtg	840
cattgggtgc	tctgtatgtgg	aatttaagcc	agacactaa	gaggaaggaa	gcagtgtga	900
aataagcagt	ggagtgggg	atagtgagag	tgaaggcctt	aacagccctg	tcaaagtgtc	960
tcgaaagcgg	aagagaatgg	tgactggaaa	tggctcttct	aaaaggaaaa	gctctagaa	1020
ggaaacgccc	tcagccacca	aacaagcaac	tagcattca	tcagaaacca	agaataacttt	1080
gagagcttc	tctgcccctc	aaaattctga	atcccaagcc	cacgttagt	gaggtggta	1140
tgacagtagt	cgcctactg	tttggtatca	tgaacttta	gaatggctt	aggagggaaa	1200
gagaagagat	gagcacagg	ggaggcctga	tcaccccgat	ttttagtcat	ctacactta	1260
tgtgcctgag	gatttctca	attctgtac	tcctggatg	aggaatgtt	ggcagattna	1320
gtctcaaac	tttgatcttg	tcatcttta	caagggtggg	aaattttatg	acgtgtacca	1380
catggatgt	cttattggag	tcagtgact	ggggctggta	ttcatgaaag	gcaactggc	1440
ccattctggc	tttcctgaaa	ttgcatttgg	cggttattca	gattccctgg	tgcagaaggg	1500
ctataaagta	gcacgagtgg	aacagactga	gactccagaa	atgtatgggg	cacgatgtag	1560
aaagatggca	catatatcca	agtatgatag	agtggtgagg	agggagatct	gtaggatcat	1620
taccaagggt	acacagactt	acagtgtgt	ggaagggtat	ccctctgaga	actacagtaa	1680
gtatctctt	agcctcaaag	aaaaagagga	agattcttct	ggccataact	gtgcatatgg	1740
tgtgtgtctt	gttgatactt	caetggaaa	gtttttcata	ggtcagttt	catagatgtc	1800
ccattgttcg	agattttagga	ctctagtgcc	acactatccc	ccagtacaag	tttttatttg	1860
aaaagggaaat	ctctcaaagg	aaactaaaac	attcttaaag	agttcattgt	cctgttct	1920
tcaggaaggt	ctgataccgg	gtcccagg	ttgggatgca	tccaaaactt	tgagaactct	1980
ccttggggaa	gaatatttt	ggggaaaggt	aagtgtggc	attgggggtga	tgttacccca	2040
ggtgcttaaa	ggatgactt	caagacttga	ttccattggg	ttgacaccag	gagagaaaag	2100
tgaatttggcc	ctctctgtct	taggtggtt	tgtcttctac	ctcaaaaaat	gccttatttg	2160
tcaggagctt	ttatcaatgg	ctaatttga	agaatataatt	cccttggatt	ctgacacagt	2220
cagcactaca	agatctgggt	ctatcttac	caaagcttat	caacgaatgg	tgcttagatgc	2280
agtgcacatta	aacaacttgg	agattttct	gaatggaaac	aatgtttcta	ctgaaggaac	2340
cctactagag	agggttgata	cttgccatac	tccttttgt	aagcggctcc	taaagcaatg	2400
gctttgtgcc	ccactctgt	accattatgc	tattaatgt	cgtctagatg	ccatagaaga	2460
cctcatgtt	gtgcctgaca	aaatctccga	agttgttagag	cttctaaaga	agcttccaga	2520
tcttggagg	ctactcgt	aaattcataa	tgttgggtct	ccccctgaaga	gtcagaacca	2580
cccagacagc	agggtataaa	tgtatgaaga	aactacatac	agcaagaaga	agattatttg	2640
ttttcttct	gtctggaaag	gatccaaatgt	aatgtgtaaa	attatagggg	tcatggaaat	2700
agttgtgtat	ggtttttaatgt	ctaaaaatct	taagcaggct	atctctctgc	agacaaaaaaa	2760
tcctgaagg	cgttttctt	atttgactgt	agaattgaac	cgatggata	cagccttta	2820
ccatggaaaag	gctcgaaaga	ctggacttat	tactccaaa	gcaggtttt	actctgatta	2880
tgaccaagct	cttgctgaca	taagagaaaa	tgaacagagc	ctcttggaa	acctagagaa	2940
acagcgaac	agaattggct	gtaggaccat	agtctatttg	gggattggta	ggaaccgtta	3000
ccagctggaa	attcctgaga	atttaccac	tcgcaatttg	ccagaagaat	acgagttgaa	3060
atctaccaag	aagggtgtt	aacgatactg	gaccaaaaact	attgaaaaga	agttggctaa	3120
tctcataaaat	gtcgaaagac	ggagggatgt	atcattgtaa	gactgtatgc	ggcgactgtt	3180
ctataactt	gataaaaaatt	acaaggactg	gcagtctgt	gtagagtgt	tcgcagttt	3240
ggatgtttt	ctgtgcctgg	ctaactatag	tcgaggggg	gatggcttct	tgtgtcggcc	3300
agtaatctt	ttggccggaa	ataccctttt	ttttagttag	ctttaaggat	cacgcctatcc	3360
ttgcattac	aaagactttt	ttggagatg	ttttttctt	aatgacattc	taataggctg	3420
tgaggaagag	gaggcggaaa	atggccaaag	stttagtgt	cttggtaact	gaccaaatat	3480
ggggggcag	tctacgctt	tgagacaggc	tggcttatta	gctgtatgg	cccagatggg	3540
ttgttacgtc	cctgtctgaa	tgtgcaggct	cacaccaatt	gataaggtgt	ttactagact	3600
ttgggcctca	gacagaataa	tgtcagggt	aagtacattt	tttggtaat	taagtgaac	3660
tgccagcata	ctcatgtcgt	caacacgaca	ttctctggtg	cttggatgt	aatttaggaag	3720

MOR-0017

PATENT

aggtaactgca acatttgatg ggacggcaat agcaaatgca gttgttaaag aacttgctga	3780
gactataaaa tgcgtacat tatttcaac tcactaccat tcatttagtag aagattattc	3840
tcaaaatgtt gctgtgcgcc taggacatat ggcatgcatt gtagaaaaatg aatgtgaaga	3900
ccccagccag gagactatta cgttccctta taaattcatt aaggagctt gtcctaaaag	3960
ctatggcttt aatgcagcaa ggcttgctaa tctcccagag gaagttattc aaaaggaca	4020
tagaaaagca agagaatttg agaagatgaa tcagtcacta cgattattc gggagttt	4080
cctggctagt gaaaggtcaa ctgttagatgc tgaagctgtc catabattgc tgactttgat	4140
taaggaatta tagactgact acattgaaag ctttgagttt acttctgaca aaggtggtaa	4200
attcagacaa cattatgatc taataaactt tatttttaa aaat	4244

**MSH3 (human) (SEQ ID NO:28)**

MSRRKPASGG LAASSSAPAR QAVLSRFFQS TGSLKSTSSS TGAADQVDPG AAAAAAPPAP	60
AFPPQLPPHV ATEIDRRKKR PLENDGPVKK KVKKVQQKEG GSDLMSGNS EPKKCLRTRN	120
VSKSLEKLKE FCCDSALPQS RVQTESLQER FAVLPKCTDF DDISLHAKN AVSSEDSKRQ	180
INQKDTTLFD LSQFGSSNTS HENLQKTASK SANKRSKSIVY TPLEQYIEM KQQHKDAVLC	240
VECGYKYRFF GEDAEIAARE LNIYCHLDHN FMTASIPTHR LFVHVRRLVA KGKVGVVVKQ	300
TETAALKAIG DNRSSLFSRK LTALYTKSTL IGEDVNPLIK LDDAVNVDEI MTDTSYILL	360
CISENKENVR DKKKGNIFIG IVGVQPATGE VVFDSFQDSA SRSELETRMS SLQPVELLLP	420
SALSEQTEAL IHRATSVSVO DDRIRVERMD NIYFEYSHAF QAVTEFYAKD TVDIKGSQII	480
SGIVNLEKPV ICSLAAIIKY LKEFNLEKML SKPENFKQLS SKMEFTING TTLRNLEILQ	540
NQTDMKTGKS LLWVLVDHTKT SFGRRLKKW VTQPLLKLRE INARLDADVSE VLHSESSVFG	600
QIENHLRKLP DIERGLCSIY HKKCSTQEFP LIVKTLHYLK SEFOAIIPAV NSHIQSDLLR	660
TVILEIPELL SPVEHYLKIL NEQAAKVGDK TELFKDLSDLF PLIKKRKDEI QGVIDEIRMH	720
LQEIRKILKN PSAQYVTVSG QEFMIEIKNS AVSCIPTDWV KGNGSTKAVSR FHSPFIVENY	780
RHLNQLREQL VLDCAEWSL FLEKFSEHYH SLCKAVHHLA TVDCIFSLAK VAKQGDYCRP	840
TVQEERKIVI KNGRHPVIDV LLGEQDQYVP NNTDLSEDSE RVMIITGPNM GGKSSYIKQV	900
ALITIMAQIG SYVPAEEATI GIVDGIFTRM GAADNIYKGR STFMEELTDT AEIIRKATSQ	960
SLVILDELGR GTSTHDGIAI AYATLEYFIR DVKSLTFLFVT HYPPVCELEK NYSHQVGNYH	1020
MGFLVSEDES KLDPGAAEQV PDFVTFLYQI TRGIAARSYG LNVAKLADVP GEILKKAHK	1080
SKELEGLEINT KRKRALKYFAK LWTMHNAQDL QKWTEEFNME ETQTSLHH	1128

**MSH3 (human DNA) (SEQ ID NO:29)**

gggcacgagc cctgccatgt ctcqccggaa gcctgcgtcg ggcggcctcg ctgcctccag	60
ctcagccct gcgaggcaag cggttttag ccgattcttc cagtcacgg gaagcctgaa	120
atccacccctt tcctccacag gtgcagccga ccaggtggac cctggcgctg cagcggccgc	180
agcgccccca gcccgcct tccgcggcca gctgcccgg cacgtagcta cagaattga	240
cagaagaaag aagagaccat tggaaaatga tggccctgtt aaaaagaaag taaagaaagt	300
ccaacaaaag gaaggaggaa gtgatctggg aatgtctggc aactctgagc caaagaaatg	360
tctgaggacc agaaatgtt caaagtctt ggaaaaattt aaagaattct gctgcattc	420
tgccttcctt caaagtagag tccagacaga atctctgcag gagagattt cagttctgcc	480
aaaatgtact gatTTGATG atatcgtct ttacacgc aagaatgcag tttcttctga	540
agattcgaaa cgtcaattt atcaaaagga cacaacactt ttgtatctca gtcagtttg	600
atcatcaaattt acaagtcatg aaaatttaca gaaaactgct tccaaatcag ctaacaaacg	660
gtccaaaagc atctatacgc cgctagaatt acaatacata gaaatgaagc agcagcacaa	720
agatgcagtt ttgtgtgtgg aatgtggata taagtataga ttcttgggg aagatgcaga	780
gattgcagcc cgagagctca atatttattt coatttagat cacaacttta tgacagcaag	840
tatacctact cacagactgt ttgttcatgt acggccctcg gtggccaaag gatataaggt	900
gggagttgtg aagcaaactg aaactgcagc attaaaggcc attggagaca acagaagtgc	960
actctttcc cgaaatttg ctgccttta tacaattt acacttattt gagaagatgt	1020
gaatccctta atcaagctgg atgatgtgtt aaatgttgat gagataatga ctgatactc	1080
taccagctat ttctgtgtca tctgtaaaaa taaggaaaat ttgtggacca aaaaaagg	1140
caacatttt attggcattt tgggagtgca gcctgccaca ggcgggttg ttgttgatag	1200
tttccaggac tctgcttctc gttcagagct agaaacccgg atgtcaagcc tgcagccagt	1260
agagctgctg ctcccttcgg ctttgcgg gcaaacagag ggcgtcatcc acagagccac	1320
atctgttagt gtgcaggatg acagaattcg agtgcggaaagg atggataaca ttatTTGA	1380
atacagccat gcttccagg cagttacaga gttttatgca aaagatacag ttgacatcaa	1440
aggttctcaa atttttctg gcattgttaa cttagagaag cctgtgattt gctcttgc	1500
tgccatcata aaatacctca aagaattcaa ctggaaaag atgctctcca aacctgagaa	1560

MOR-0017

PATENT

ttttaaacag	ctatcaagta	aatatggaaatt	tatgacaatt	aatggaacaa	cattaaggaa	1620
tctggaaatc	ctacagaatc	agactgatata	aaaaacccaaa	ggaagtttgc	tgtggggtt	1680
agaccacact	aaaacttcat	ttggagagacg	gaagttaaag	aagtgggtga	cccagccact	1740
ccttaaatta	agggaaataa	atgcccggct	tgtatgtt	tcggaagttc	tccattcaga	1800
atcttagtgt	tttggtcaga	tagaaaatca	tctacgtaaa	ttgcccgaca	tagagagggg	1860
actctgttagc	atttatcaca	aaaaatgttc	tacccaaagag	ttcttcttga	ttgtcaaaac	1920
tttatatcac	ctaaagtccg	aatttcaagc	aataataatcc	gctgttaatt	cccacattca	1980
gtcagacttg	ctccggaccg	ttatTTTGA	aatttctgaa	ctcctcagtc	cagtggagca	2040
ttacttaaag	atactcaatg	aacaagttgc	caaagtggg	gataaaaactg	aatttattaa	2100
agacctttct	gacttccctt	taataaaaaa	gaggaaggat	gaaattcaag	gtgttattga	2160
cgagatccga	atgcatttgc	aagaaatacg	aaaaataacta	aaaaatcctt	ctgcacaata	2220
tgtgacagta	tcaggacagg	agtttatgtat	agaaaataaaag	aactctgctg	tatcttgtat	2280
accaactgtat	ttggtaaagg	ttggaagcac	aaaagctgt	agccgcttc	actctcctt	2340
tattgtagaa	aattacagac	atctgaatca	gtcggggag	cagctagtcc	ttgactgcag	2400
tgctgaatgg	cttgattttc	tagagaaaatt	cagtgaacat	tatcactcct	tgtgtaaagc	2460
agtgcacatc	ctagcaactg	ttgactgtat	tttctccctg	gccaaggctcg	ctaagcaagg	2520
agattactgc	agaccaactg	tacaagaaga	aagaaaaatt	gtaataaaaa	atggaagcca	2580
ccctgtgatt	gatgtgttgc	tgggagaaca	ggatcaat	gtcccaaata	atacagattt	2640
atcagaggac	tcagagagag	taatgataat	tacccggacca	aacatgggtg	gaaagagactc	2700
ctacataaaa	caagtgcat	tgattaccat	catggctca	attggctct	atgttctgc	2760
agaagaagcg	acaattggga	tttggtatgg	cattttcaca	aggatgggt	ctgcagacaa	2820
tatataaaa	ggacggagta	catttatgga	agaactgact	gacacagcag	aaataatcag	2880
aaaagcaaca	tcacagtct	ttggttatctt	ggatgaacta	ggaagaggg	cgagcactca	2940
tgatggaaatt	gcattgcct	atgctacact	tgagtatttc	atcagagatg	tgaatcctt	3000
aaccctgttt	gtcacccatt	atccggcagt	ttgtgaacta	aaaaaaaatt	actcacacca	3060
ggtgggaaat	taccacatgg	gattcttgg	cagtgaggat	gaaagcaaac	ttgatccagg	3120
cgcagcagaa	caagtccctg	attttgcac	cttcctttac	caaataacta	gaggaattgc	3180
agcaaggagt	tatggattaa	atgtggctaa	actagcagat	gttcctggag	aaattttgaa	3240
gaaaggcact	cacaagtcaa	aagagctgga	aggattataa	aatacggaaa	gaaagagact	3300
caagtatTTT	gcaaaagtat	ggacgtatca	taatgcacaa	gacctgcaga	agtggacaga	3360
ggagttcac	atggaagaaaa	cacagactt	totttctcat	taaaaatgaa	actacatttgc	3420
tgaacaaaaaa	atggagaattt	aaaatatacc	actgtacaaa	ataactctcc	actaacaggc	3480
tatctttgt	tgacatgtat	gcataaaaatt	atgaccatgg	tatattctta	ttggaaacag	3540
agaggttttt	ctgaagacag	tcttttcaa	gtttctgtct	tcctaacttt	tctacgtata	3600
aacactcttgc	aatagacttc	cactttgtaa	ttagaaaatt	ttatggacag	taagtccagt	3660
aaaggccttaa	gtggcagaat	ataattccca	agcttttgg	gggtgatata	aaaatttact	3720
tgatattttt	atttggttca	gttcagataa	ttggcaactg	ggtgaatctg	gcaggaatct	3780
atccattgaa	ctaaaataat	tttattatgc	aaccagttt	tccaccaaga	acataagaat	3840
tttttataag	tagaaagaat	tggccaggca	tgttggctca	tgcctgtat	cccagcactt	3900
tgggaggcca	aggtaggcag	atcacctgag	gtcaggaggt	caagaccagc	ctggccaaaca	3960
tggcaaaacc	ccatctttac	aaaaatata	aagtacatct	ctactaaaa	tacgaaaaaa	4020
tttagctggc	atgtggcgc	acacctgtat	tcccgactac	tccggaggct	gaggcaggag	4080
aatctcttgc	acctgggggg	cggagggtgc	aatgagccga	gatcacgtca	ctgcactc	4140
gcttggccaa	cagagcaaga	ctccatctca	aaaaagaaaa	aagaaaaagaa	atagaattat	4200
caagctttta	aaaactagag	cacagaagga	ataaggtcat	gaaatttaaa	aggttaaata	4260
ttgtcatagg	attaaggcgt	ttaaagattt	ttggatgaaa	ttatttgtca	ttcattcaag	4320
taataaaat	ttaatgaata	cttgctataa	aaaaaaaaaa	aaaaaaaaaa	aaaa	4374

Each reference cited herein is hereby incorporated by reference in its entirety.